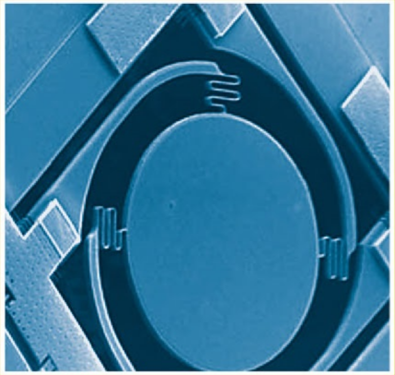
WOODHEAD PUBLISHING SERIES IN ELECTRONIC AND OPTICAL MATERIALS



Smart sensors and MEMS

Intelligent devices and microsystems for industrial applications

Edited by Stoyan Nihtianov and Antonio Luque

WP
WOODHEAD
PUBLISHING

VISIT…

LANZAROTE
Caliente.COM

Smart sensors and MEMS

Related titles:

MEMS for automotive and aerospace applications
(ISBN 978-0-85709-118-5)

Handbook of MEMS for wireless and mobile applications
(ISBN 978-0-85709-271-7)

MEMS for biomedical applications
(ISBN 978-0-85709-129-1)

Details of these books and a complete list of titles from Woodhead Publishing can be obtained by:

- visiting our web site at www.woodheadpublishing.com
- contacting Customer Services (e-mail: sales@woodheadpublishing.com; fax: +44 (0) 1223 832819; tel.: +44 (0) 1223 499140 ext. 130; address: Woodhead Publishing Limited, 80 High Street, Sawston, Cambridge CB22 3HJ, UK)
- in North America, contacting our US office (e-mail: usmarketing@woodheadpublishing.com; tel.: (215) 928 9112; address: Woodhead Publishing, 1518 Walnut Street, Suite 1100, Philadelphia, PA 19102-3406, USA)

If you would like e-versions of our content, please visit our online platform: www.woodheadpublishingonline.com. Please recommend it to your librarian so that everyone in your institution can benefit from the wealth of content on the site.

We are always happy to receive suggestions for new books from potential editors. To enquire about contributing to our Electronic and Optical Materials series, please send your name, contact address and details of the topic/s you are interested in to laura.pugh@woodheadpublishing.com. We look forward to hearing from you.

The team responsible for publishing this book:

Commissioning Editor: Laura Pugh
Publications Coordinator: Emily Cole
Project Editor: Anneka Hess
Editorial and Production Manager: Mary Campbell
Production Editor: Adam Hooper
Project Manager: Newgen Knowledge Works Pvt Ltd
Copyeditor: Newgen Knowledge Works Pvt Ltd
Proofreader: Newgen Knowledge Works Pvt Ltd
Cover Designer: Terry Callanan

Woodhead Publishing Series in Electronic and Optical Materials:
Number 51

Smart sensors and MEMS

Intelligent devices and microsystems for industrial applications

---

Edited by
Stoyan Nihtianov and
Antonio Luque

WP
WOODHEAD
PUBLISHING

Oxford Cambridge Philadelphia New Delhi

Published by Woodhead Publishing Limited,
80 High Street, Sawston, Cambridge CB22 3HJ, UK
www.woodheadpublishing.com
www.woodheadpublishingonline.com

Woodhead Publishing, 1518 Walnut Street, Suite 1100, Philadelphia,
PA 19102-3406, USA

Woodhead Publishing India Private Limited, 303 Vardaan House, 7/28 Ansari Road,
Daryaganj, New Delhi – 110002, India
www.woodheadpublishingindia.com

First published 2014, Woodhead Publishing Limited
© Woodhead Publishing Limited, 2014. The publisher has made every effort to ensure that permission for copyright material has been obtained by authors wishing to use such material. The authors and the publisher will be glad to hear from any copyright holder it has not been possible to contact.
The authors have asserted their moral rights.

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. Reasonable efforts have been made to publish reliable data and information, but the authors and the publishers cannot assume responsibility for the validity of all materials. Neither the authors nor the publishers, nor anyone else associated with this publication, shall be liable for any loss, damage or liability directly or indirectly caused or alleged to be caused by this book.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming and recording, or by any information storage or retrieval system, without permission in writing from Woodhead Publishing Limited.

The consent of Woodhead Publishing Limited does not extend to copying for general distribution, for promotion, for creating new works, or for resale. Specific permission must be obtained in writing from Woodhead Publishing Limited for such copying.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation, without intent to infringe.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

Library of Congress Control Number: 2013948530

ISBN 978-0-85709-502-2 (print)
ISBN 978-0-85709-929-7 (online)
ISSN 2050-1501 Woodhead Publishing Series in Electronic and Optical Materials (print)
ISSN 2050-151X Woodhead Publishing Series in Electronic and Optical Materials (online)

The publisher's policy is to use permanent paper from mills that operate a sustainable forestry policy, and which has been manufactured from pulp which is processed using acid-free and elemental chlorine-free practices. Furthermore, the publisher ensures that the text paper and cover board used have met acceptable environmental accreditation standards.

Typeset by Newgen Knowledge Works Pvt Ltd
Printed by Lightning Source

Contents

Contributor contact details xiii
Woodhead Publishing Series in Electronic and Optical Materials xvii
Preface xxi

Part I Smart sensors for industrial applications 1

1 What makes sensor devices and microsystems 'intelligent' or 'smart'? 3
R. Taymanov and K. Sapozhnikova, D. I. Mendeleyev Institute for Metrology, Russia

1.1 Introduction 3
1.2 Interpretation of terms related to sensors 5
1.3 Key trends in the development of sensors (sensor devices) and MEMS 10
1.4 Suggestions for improving terminology in the field 17
1.5 Conclusion 21
1.6 Acknowledgment 21
1.7 References 21

2 Direct interface circuits for sensors 27
F. Reverter, Universitat Politècnica de Catalunya (UPC), Spain

2.1 Introduction 27
2.2 Sensors 29
2.3 Microcontrollers 34
2.4 Interface circuits 40
2.5 Applications 53
2.6 Future trends 58

2.7 Sources of further information and advice 59
2.8 References 60

3 Capacitive sensors for displacement measurement in the sub-nanometer range 63
S. Xia, NXP Semiconductors, The Netherlands and S. Nihtianov, Delft University of Technology, The Netherlands

3.1 Introduction 63
3.2 Challenges for sub-nanometer displacement measurement with capacitive sensors 63
3.3 Offset capacitance cancellation technique 65
3.4 Capacitance-to-digital converter (CDC) with offset capacitance cancellation and calibration functions 67
3.5 Conclusion 74
3.6 References 74

4 Integrated inductive displacement sensors for harsh industrial environments 76
M. R. Nabavi, Catena Microelectronics BV, The Netherlands and S. Nihtianov, Delft University of Technology, The Netherlands

4.1 Introduction 76
4.2 Principles of operation and practical limitations for eddy-current sensors (ECSs) 78
4.3 Design requirements in precision industrial applications 84
4.4 State-of-the-art ECS interfaces 86
4.5 ECS interfaces with an LC oscillator and ratiometric measurement 89
4.6 Conclusion 99
4.7 References 100

5 Advanced silicon radiation detectors in the vacuum ultraviolet (VUV) and the extreme ultraviolet (EUV) spectral range 102
A. Gottwald and F. Scholze, Physikalisch-Technische Bundesanstalt, Germany

5.1 Introductory overview 102
5.2 Challenges for radiation detection in the EUV and VUV spectral ranges 103

5.3 Device solutions for radiation detection in the EUV and VUV spectral ranges 107
5.4 Methods of radiometric investigation and characterization 111
5.5 Spectral responsivity and radiation hardness of EUV and VUV radiation detectors 113
5.6 Future trends 120
5.7 References 120

6 Integrated polarization analyzing CMOS image sensors for detection and signal processing 124
M. Sarkar, Indian Institute of Technology, India and
A. J. P. Theuwissen, Delft University of Technology,
The Netherlands

6.1 Introduction 124
6.2 Polarization vision 127
6.3 Polarization cameras 135
6.4 Design of a polarization sensor 141
6.5 Polarization vision in machine vision applications: material classification 143
6.6 Conclusion 149
6.7 References 150

7 Advanced interfaces for resistive sensors 153
A. Flammini and A. Depari, University of Brescia, Italy

7.1 Introduction 153
7.2 Resistive sensors 154
7.3 Voltamperometric resistance estimation 160
7.4 Resistance-to-time conversion methods 173
7.5 Conclusion and future trends 201
7.6 References 202

8 Reconfigurable ultrasonic smart sensor platform for nondestructive evaluation and imaging applications 205
J. Saniie, E. Oruklu and S. Gilliland, Illinois Institute of Technology, USA and S. Aslan, Texas State University, USA

8.1 Introduction 205
8.2 Fundamentals of ultrasonic sensing and pulse-echo measurements 207
8.3 Reconfigurable ultrasonic smart sensor platform (RUSSP) design 209

8.4 Algorithms used in evaluation of RUSSP 215
8.5 Hardware realization of ultrasonic imaging algorithms using RUSSP 219
8.6 Future trends 226
8.7 Conclusion 227
8.8 Sources of further information and advice 227
8.9 References 227

9 Advanced optical incremental sensors: encoders and interferometers 230
S. J. A. G. Cosijns and M. J. Jansen, ASML, The Netherlands

9.1 Introduction 230
9.2 Displacement interferometers 231
9.3 Sources of error and compensation methods 243
9.4 Optical encoders 249
9.5 Design considerations 267
9.6 Current and future trends 273
9.7 Conclusion 274
9.8 References 274

Part II Smart micro-electro-mechanical systems (MEMS) for industrial applications 279

10 Microfabrication technologies used for creating smart devices for industrial applications 281
J. M. Quero, F. Perdigones and C. Aracil, University of Seville, Spain

10.1 Introduction 281
10.2 MEMS design and modeling 282
10.3 Materials 283
10.4 Microfabrication processes 287
10.5 Simulation 297
10.6 Conclusion 303
10.7 References 303

11 Microactuators: design and technology 305
L. Li and Z. J. Chew, Swansea University, UK

11.1 Introduction 305
11.2 Driving principles for actuators 307

11.3 Design and analysis of microactuators 317
11.4 Conclusion 336
11.5 References 337

12 Dynamic behavior of smart MEMS in industrial applications 349
M. Pustan, C. Birleanu and C. Dudescu, Technical University of Cluj-Napoca, Romania and J-C. Golinval, University of Liège, Belgium

12.1 Introduction 349
12.2 Resonant frequency response of smart MEMS vibrating structures 350
12.3 Quality factor and the loss coefficient of smart MEMS vibrating structures 358
12.4 Industrial applications: resonant accelerometers 364
12.5 References 364

13 MEMS integrating motion and displacement sensors 366
G. Langfelder and A. Tocchio, Politecnico di Milano, Italy

13.1 Introduction 366
13.2 Technical description of MEMS motion sensors: MEMS accelerometer 369
13.3 MEMS gyroscope 380
13.4 MEMS magnetometer 390
13.5 Conclusion and future trends 395
13.6 References 397

14 MEMS print heads for industrial printing 402
S. Lee and J. Choi, Sungkyunkwan University (SKKU), Republic of Korea

14.1 Introduction 402
14.2 Electro-hydro-dynamics (EHD) print head droplet ejection 407
14.3 EHD smart printing system 413
14.4 Case study: EHD printing applications 423
14.5 Conclusion 427
14.6 References 428

15 Photovoltaic and fuel cells in power MEMS for smart energy management 431
J. García and F. J. Delgado, University of Seville, Spain and P. Ortega and S. Bermejo, Polytechnic University of Catalonia, Spain

15.1 Introduction 431
15.2 Photovoltaic mini-generators 433
15.3 Applications of photovoltaic mini-generators 445
15.4 Micro-fuel cells 449
15.5 Applications of micro-fuel cells 456
15.6 Smart energy management with sun sensors 457
15.7 Conclusion 468
15.8 References 468

16 Radio frequency (RF)-MEMS for smart communication microsystems 472
D. Dubuc and K. Grenier, LAAS-CNRS, University of Toulouse, France

16.1 Introduction 472
16.2 RF-MEMS technology and devices 473
16.3 RF-MEMS-based circuits for smart communication microsystems 478
16.4 RF-MEMS reliability 482
16.5 RF-MEMS power capability 483
16.6 Co-integration of RF-MEMS-based circuits with integrated circuits (IC) 485
16.7 Conclusion 487
16.8 References 488

17 Smart acoustic sensor array (SASA) system for real-time sound processing applications 492
M. Turqueti, E. Oruklu and J. Saniie, Illinois Institute of Technology, USA

17.1 Introduction 492
17.2 MEMS microphones 494
17.3 Fundamentals of acoustic sensor arrays and applications 495
17.4 Design and implementation of a smart acoustic MEMS array (AMA) 496

17.5 System implementation of AMA and compact and programmable data acquisition nodes (CAPTAN) 500
17.6 SASA system operation 502
17.7 SASA system calibration 503
17.8 Sensor array for time-of-flight measurements 505
17.9 3D sound source localization 511
17.10 SASA system for mapping of the heart sound 512
17.11 Conclusion 515
17.12 References 515

Index *518*

Contributor contact details

(* = main contact)

Editors

S. Nihtianov
Delft University of Technology
Mekelweg 4
2628CD Delft, The Netherlands

E-mail: S.Nihtianov@tudelft.nl

A. Luque
Universidad de Sevilla
Dpto. de Ingeniería Electrónica
Escuela Técnica Superior de Ingeniería
Camino de los Descubrimientos s/n
41092 Seville, Spain

E-mail: aluque@gte.esi.us.es

Chapter 1

R. Taymanov* and K. Sapozhnikova
D.I. Mendeleyev Institute for Metrology
19 Moskovsky pr
St.Petersburg, 190005, Russia

E-mail: taymanov@vniim.ru; kseniavs@yandex.ru

Chapter 2

F. Reverter
Department of Electronic Engineering
Universitat Politècnica de Catalunya (UPC)
C/. Esteve Terradas, 7, Building C4
08860, Castelldefels, Barcelona, Spain

E-mail: ferran.reverter@upc.edu

Chapter 3

S. Xia*
NXP Semiconductors
High Tech Campus Eindhoven 46
5656 AG Eindhoven, The Netherlands

E-mail:sha.xia@nxp.com

S. Nihtianov
Mekelweg 4
Delft University of Technology
2628CD Delft, The Netherlands

E-mail: sha.xia@nxp.com

Chapter 4

M. R. Nabavi*
Catena Microelectronics BV
Elektronicaweg 40
2628 XG Delft, The Netherlands

E-mail: mnabavi@catena.nl

S. Nihtianov
Delft University of Technology
Mekelweg 4
2628CD Delft, The Netherlands

E-mail: S.Nihtianov@tudelft.nl

Chapter 5

F. Scholze* and A. Gottwald
Physikalisch-Technische Bundesanstalt
Abbestraße 2-12
10587, Berlin, Germany

E-mail: frank.scholze@ptb.de

Chapter 6

M. Sarkar*
Electrical Engineering Department
Indian Institute of Technology Delhi
Hauz Khas 110016, New Delhi, India

E-mail: msarkar@ee.iitd.ac.in

A. J. P. Theuwissen
Electronic Instrumentation Laboratory,
Delft University of Technology
2628CD Delft, The Netherlands

E-mail: a.theuwissen@scarlet.be.

Chapter 7

A. Flammini* and A. Depari
Department of Information Engineering
University of Brescia
Via Branze 38
25123, Brescia, Italy

E-mail: alessandra.flammini@ing.unibs.it

Chapter 8

J. Saniie
Department of Electrical and Computer Engineering
Illinois Institute of Technology
3301 South Dearborn Street
Chicago, Illinois 60616, USA

E-mail: sansonic@ece.iit.edu

Chapter 9

S. J. A. G. Cosijns and M. J. Jansen*
ASML Netherlands B.V
De Run 6501
5504 DR, Veldhoven,The Netherlands

E-mail: Maarten.Jansen@asml.com; Suzanne.Cosijns@asml.com

Chapter 10

J. M. Quero*, F. Perdigones and C. Aracil
Universidad de Sevilla
Dpto. de Ingeniería Electrónica
Escuela Técnica Superior de Ingeniería
Camino de los Descubrimientos s/n
41092 Seville, Spain

E-mail: quero@us.es

Chapter 11

L. Li* and Z. J. Chew
College of Engineering
Swansea University
Swansea, Wales, UK

E-mail: L.Li@swansea.ac.uk

Chapter 12

M. Pustan*, C. Birleanu and C. Dudescu
Technical University of Cluj-Napoca
Memorandumului Street 28
Cluj-Napoca, 400114, Romania

E-mail: Marius.Pustan@omt.utcluj.ro

J-C. Golinval
University of Liège
Chemin des Chevreuils 1, 4000, Liège, Belgium

E-mail: JC.Golinval@ulg.ac.be

Chapter 13

G. Langfelder* and A. Tocchio
Dipartimento di Elettronica e Informazione
Politecnico di Milano
via Golgi 40
20133, Milano, Italy

E-mail: langfelder@elet.polimi.it

Chapter 14

S. Lee* and J. Choi
Department of Electrical and Computer Engineering
Sungkyunkwan University (SKKU)
83646-ho
300, Chunchun-dong
Jangan-gu, Suwon
Kyunggi-do, 440-746
Republic of Korea

E-mail: lsh@ece.skku.ac.kr; yong.steinkopf@gmail.com

Chapter 15

J. García* and F. J. Delgado
Universidad de Sevilla
Dpto. de Ingeniería Electrónica
Escuela Técnica Superior de Ingeniería
Camino de los Descubrimientos s/n
41092 Seville, Spain

E-mail: juangarcia@us.es; fjdelgado@gte.esi.us.es

P. Ortega and S. Bermejo
Department of Electronic Engineering, Micro and Nanotechnologies Group MNT
UniversitatPolitècnica de Catalunya (UPC)
C/ Jordi Girona 1-3, Building C-4
Barcelona 08034, Spain

E-mail: pablo.rafael.ortega@upc.edu; sandra.bermejo@upc.edu

Chapter 16

D. Dubuc* and K. Grenier
LAAS-CNRS, University de Toulouse
7 Avenue du Colonel Roche
F-31400 Toulouse, France

E-mail: dubuc@laas.fr; grenier@laas.fr

Chapter 17

M. Turqueti, E. Oruklu and J. Saniie
Department of Electrical and Computer Engineering
Illinois Institute of Technology
3301 South Dearborn Street
Chicago, Illinois 60616, USA

E-mail: sansonic@ece.iit.edu

Woodhead Publishing Series in Electronic and Optical Materials

1 **Circuit analysis**
J. E. Whitehouse
2 **Signal processing in electronic communications: For engineers and mathematicians**
M. J. Chapman, D. P. Goodall and N. C. Steele
3 **Pattern recognition and image processing**
D. Luo
4 **Digital filters and signal processing in electronic engineering: Theory, applications, architecture, code**
S. M. Bozic and R. J. Chance
5 **Cable engineering for local area networks**
B. J. Elliott
6 **Designing a structured cabling system to ISO 11801: Cross-referenced to European CENELEC and American Standards, Second Edition**
B. J. Elliott
7 **Microscopy techniques for materials science**
A. Clarke and C. Eberhardt
8 **Materials for energy conversion devices**
Edited by C. C. Sorrell, J. Nowotny and S. Sugihara
9 **Digital image processing: Mathematical and computational methods, Second Edition**
J. M. Blackledge
10 **Nanolithography and patterning techniques in microelectronics**
Edited by D. Bucknall
11 **Digital signal processing: Mathematical and computational methods, software development and applications, Second Edition**
J. M. Blackledge
12 **Handbook of advanced dielectric, piezoelectric and ferroelectric materials: Synthesis, properties and applications**
Edited by Z.-G. Ye
13 **Materials for fuel cells**
Edited by M. Gasik

14 **Solid-state hydrogen storage: Materials and chemistry**
Edited by G. Walker
15 **Laser cooling of solids**
S. V. Petrushkin and V. V. Samartsev
16 **Polymer electrolytes: Fundamentals and applications**
Edited by C. A. C. Sequeira and D. A. F. Santos
17 **Advanced piezoelectric materials: Science and technology**
Edited by K. Uchino
18 **Optical switches: Materials and design**
Edited by S. J. Chua and B. Li
19 **Advanced adhesives in electronics: Materials, properties and applications**
Edited by M. O. Alam and C. Bailey
20 **Thin film growth: Physics, materials science and applications**
Edited by Z. Cao
21 **Electromigration in thin films and electronic devices: Materials and reliability**
Edited by C.-U. Kim
22 **In situ characterization of thin film growth**
Edited by G. Koster and G. Rijnders
23 **Silicon-germanium (SiGe) nanostructures: Production, properties and applications in electronics**
Edited by Y. Shiraki and N. Usami
24 **High-temperature superconductors**
Edited by X. G. Qiu
25 **Introduction to the physics of nanoelectronics**
S. G. Tan and M. B. A. Jalil
26 **Printed films: Materials science and applications in sensors, electronics and photonics**
Edited by M. Prudenziati and J. Hormadaly
27 **Laser growth and processing of photonic devices**
Edited by N. A. Vainos
28 **Quantum optics with semiconductor nanostructures**
Edited by F. Jahnke
29 **Ultrasonic transducers: Materials and design for sensors, actuators and medical applications**
Edited by K. Nakamura
30 **Waste electrical and electronic equipment (WEEE) handbook**
Edited by V. Goodship and A. Stevels
31 **Applications of ATILA FEM software to smart materials: Case studies in designing devices**
Edited by K. Uchino and J.-C. Debus

32 **MEMS for automotive and aerospace applications**
Edited by M. Kraft and N. M. White
33 **Semiconductor lasers: Fundamentals and applications**
Edited by A. Baranov and E. Tournie
34 **Handbook of terahertz technology for imaging, sensing and communications**
Edited by D. Saeedkia
35 **Handbook of solid-state lasers: Materials, systems and applications**
Edited by B. Denker and E. Shklovsky
36 **Organic light-emitting diodes (OLEDs): Materials, devices and applications**
Edited by A. Buckley
37 **Lasers for medical applications: Diagnostics, therapy and surgery**
Edited by H. Jelínková
38 **Semiconductor gas sensors**
Edited by R. Jaaniso and O. K. Tan
39 **Handbook of organic materials for optical and (opto) electronic devices: Properties and applications**
Edited by O. Ostroverkhova
40 **Metallic films for electronic, optical and magnetic applications: Structure, processing and properties**
Edited by K. Barmak and K. Coffey
41 **Handbook of laser welding technologies**
Edited by S. Katayama
42 **Nanolithography: The art of fabricating nanoelectronic and nanophotonic devices and systems**
Edited by M. Feldman
43 **Laser spectroscopy for sensing: Fundamentals, techniques and applications**
Edited by M. Baudelet
44 **Chalcogenide glasses: Preparation, properties and applications**
Edited by J.-L. Adam and X. Zhang
45 **Handbook of MEMS for wireless and mobile applications**
Edited by D. Uttamchandani
46 **Subsea optics and imaging**
Edited by J. Watson and O. Zielinski
47 **Carbon nanotubes and graphene for photonic applications**
Edited by S. Yamashita, Y. Saito and J. H. Choi
48 **Optical biomimetics: Materials and applications**
Edited by M. Large
49 **Optical thin films and coatings**
Edited by A. Piegari and F. Flory

50 **Computer design of diffractive optics**
Edited by V. A. Soifer
51 **Smart sensors and MEMS: Intelligent devices and microsystems for industrial applications**
Edited by S. Nihtianov and A. Luque
52 **Fundamentals of femtosecond optics**
S. A. Kozlov and V. V. Samartsev
53 **Nanostructured semiconductor oxides for the next generation of electronics and functional devices: Properties and applications**
S. Zhuiykov
54 **Nitride semiconductor light-emitting diodes (LEDs): Materials, performance and applications**
Edited by J. J. Huang, H. C. Kuo and S. C. Shen
55 **Sensor technologies for civil infrastructures**
Volume 1: Sensing hardware and data collection for performance assessment
Edited by M. Wang, J. Lynch and H. Sohn
56 **Sensor technologies for civil infrastructures**
Volume 2: Applications in structural health monitoring
Edited by M. Wang, J. Lynch and H. Sohn
57 **Graphene: Properties, preparation, characterisation and devices**
Edited by V. Skákalová and A. B. Kaiser
58 **Handbook of silicon-on-insulator (SOI) technology**
Edited by O. Kononchuk and B.-Y. Nguyen
59 **Biological identification: DNA amplification and sequencing, optical sensing, lab-on-chip and portable systems**
Edited by P. Schaudies
60 **High performance silicon imaging: Fundamentals and applications of CMOS and CCD sensors**
Edited by D. Durini
61 **Nanosensors for chemical and biological applications: Sensing with nanotubes, nanowires and nanoparticles**
Edited by K. C. Honeychurch
62 **Composite magnetoelectrics: Materials, structures, and applications**
G. Srinivasan, S. Priya, and N. Sun
63 **Quantum information processing with diamond: Principles and applications**
Edited by S. Prawer and I. Aharonovich
64 **Advances in nonvolatile memory and storage technology**
Edited by Y. Nishi
65 **Laser surface engineering: Processes and applications**
Edited by J. Lawrence, C. Dowding, D. Waugh and J. Griffiths

Preface

The progress in micro- and nanotechnology provides powerful means for creating sophisticated sensor devices and systems which can operate properly and reliably in a wide range of working conditions. Such sensor solutions are an essential part of modern industrial equipment. The proper functioning of a machine, just like a living organism, relies on the information from the surrounding environment. This information is used to make the right decisions and to tune the operating parameters for delivering the best product quality in the most efficient way. The sensors are the providers of this crucial information. The performance of an industrial machine cannot be better than the performance of its sensors. That is why the progress in the sensor technology directly affects the productivity and the quality of industrial products.

The book covers a collection of key topics in the field of sensors and micro-electro-mechanical structures (MEMS). Despite the numerous efforts in the last few decades to put order in the terminology and definitions in the field of sensors, still there is considerable ambiguity about what is a 'sensor' and when it is 'smart'. An original interpretation of the 'smartness' and 'intelligence' of sensors and MEMS is offered in Chapter 1. Chapters 2–9 are dedicated to advanced sensor solutions classified with respect to the phenomenon used to convert the physical quantity to electrical signal: capacitive, inductive (eddy current), resistive, optical, ultrasound, incremental sensors. Significant attention is given to the interface electronics, as an important factor for the sensor performance and especially for its 'intelligence'.

Microsystems or MEMS have particular fabrication processes and techniques, nowadays very standardized in some fields, but still with major improvements and research being made in others, such as the use of novel materials. All the fabrication processes are described in detail in Chapter 10. These techniques will be used in the following chapters, where specific devices and applications are described, for example motion sensors, printer heads, photovoltaic and fuel cells, communication devices, or acoustic sensors, covered in chapters 13–17. In all cases, the industrial applications of the particular components are included. The dynamic behavior of these devices is the subject of Chapter 12. Finally, a book on sensors and MEMS cannot be complete without at least a brief description of the complement of sensors, the actuators, which are described in Chapter 11.

S. Nihtianov
A. Luque

1

What makes sensor devices and microsystems ‘intelligent’ or ‘smart’?

R. TAYMANOV and K. SAPOZHNIKOVA,
D. I. Mendeleyev Institute for Metrology, Russia

DOI: 10.1533/9780857099297.1.3

Abstract: This chapter demonstrates the ambiguity of many terms and definitions applied to modern sensors and micro-electro-mechanical systems, including those with artificial intelligence. This ambiguity complicates communication between specialists, makes the search for required information more difficult and becomes, in the long run, an obstacle to the development of metrology and sensorics. The authors proposed a new approach that enables the terminology to be improved. It relies on an analogy between biological and technical evolution. Reasoning from this approach, the authors have formed a group of main interrelated terms in the field under consideration.

Key words: sensor, sensor device, intelligent sensor device, smart sensor device, metrological self-check.

1.1 Introduction

Development of science and associated technical terminology is an integrated process. As Kuhn (1962) has shown, scientific progress is a series of phases of ‘scientific revolutions’ related to spasmodic changes of concepts (paradigm shifts) concerning the further development of sciences. This is characteristic not only of fundamental sciences, but also for applied sciences – in particular, the sciences associated with the development of control systems and measuring instruments, including sensors and micro-electro-mechanical systems (MEMS).

The ‘paradigm shifts’ relate to terminology, too. Changes in this domain are caused by a desire to define the notions which are used for new concepts. However, the development of the applied sciences is rigidly connected with the market: scientists and designers create products for the market; the market, in its turn, stimulates and invests in the development of new products. This interaction can influence the terminology. Manufactures and dealers tend to advertise their goods by ‘embellishing’ the attributes of these goods. Without sufficient information about the quality of these goods, customers make their choice based on the name of the products, which in some cases is not adequate. This may contradict the interests of the society as a whole.

Lack of monosemantic terminology, especially under conditions of immense technological progress, as well as economic globalization, hinders communication between specialists and the search for necessary information – in the long run, becoming an obstacle to the development of science and engineering. The ambiguity of terms and their definitions leads to (and, sometimes, is a means of) dishonest price competition between some goods – which is determined, among other factors, by the names of these goods (Taymanov and Sapozhnikova, 2009a). To a great extent, the same situation pertains to sensors and MEMS.

Progress in metrology and technology has resulted in the rapid growth of the available range of integrated sensors and MEMS with new features. On the basis of such devices, technically complex set-ups and aggregated products have been developed. These configurations contain hundreds, or even thousands, of embedded sensing devices. A large number of built-in sensors and MEMS require a significant increase in their functional reliability, including metrological reliability.

The appearance of sensors and MEMS with embedded signal processing based on an analog-to-digital converter (ADC), a microcontroller and other additional components has led to new trends in the provision of metrological traceability (Tarbeyev *et al.*, 2007). These trends relate to the new functions of the devices under consideration and, consequently, the development of new metrological procedures, which should be specified in technical documentation. Corresponding theoretical and applied papers have become considerably more prolific over the last decade (Taymanov and Sapozhnikova, 2010b; Tarbeyev *et al.*, 2007). It is therefore becoming increasingly important that the terminology in the field of sensors and MEMS, including those with artificial intelligence, is harmonized. It is believed that many living creatures have intelligence, but crows, cats, dogs and humans are distinguished by different levels of intelligence. In a similar way, modern sensors and MEMS also have different levels of intelligence. It is necessary to find a universal criterion for determining whether these devices have intelligence and, if so, its level.

One would think that a new concept can be defined with the help of an existing term supplemented by a set of qualifying adjectives. However, if the number of additional adjectives for new concepts which are used frequently is greater than two, a new term will inevitably appear. The fact is that in scientific and technical papers, books, and in advertisement literature in particular, some of the supplemental attributes will be voluntarily or involuntarily omitted. Depending on what specific words have been excluded from the text and the level of the reader's experience, the text can be interpreted in a variety of ways. The situation redoubles, since some terms defined many years ago, have lost their unambiguity due to the development of technology (Taymanov and Sapozhnikova, 2009b). The rate at which terminology vocabularies are updated, including the International Vocabulary of Metrology (VIM, 2012), as well as terms given in normative documents, lags behind the rate at which new terms appear. Names for new concepts are born and spread in numerous scientific publications. As a result, it is quite often possible to come across terms which are interpreted differently. On the other hand, in some cases different terms are used for similar concepts.

1.2 Interpretation of terms related to sensors

In recent years, the intensive development of measurement science and sensorics has resulted in implementing sensors with new functions. And here a question arises with respect to their names: whether or not they reflect the special features of the sensors with new functions.

1.2.1 About the term 'sensor'

Interpretation of this term in various dictionaries and normative documents varies.

According to the Institute of Electrical and Electronics Engineers Standard (IEEE, 1999), 'sensor' means a 'component providing a useful output in response to a physical, chemical, or biological phenomenon. This component may already have some signal conditioning associated with it'. In the same document, an explanation is given that the signal conditioning 'involves operations such as amplification, compensation, filtering, and normalization'. The document continues: 'Examples: platinum resistance temperature detector, humidity sensor with voltage output, light sensor with frequency output, pH probe, and piezoresistive bridge'. It follows from this definition that two versions of a sensor are possible: a sensor with internal signal conditioning or without it.

In the IEEE Standard published subsequently (IEEE, 2003), the term 'sensor' means 'a transducer that converts a physical, biological, or chemical parameter into an electrical signal'. (*Note*: According to IEEE Standard (IEEE, 1999), 'transducer' means 'a device converting energy from one domain into another, calibrated to minimize the errors in the conversion process'.)

According to (VIM, 2012), 'measuring transducer' means a 'device, used in measurement, that provides an output quantity having a specified relation to the input quantity'. In the Russian recommendation (RMG 29 1999 (2000)), the term '*datchik*' ('giver') has been introduced. It is defined as a self-contained primary transducer that generates measurement signals at its output (it 'gives' information). In the note to this term, it is emphasized that the '*datchik*' can be placed at a distance with regard to a transducer receiving its signals. According to NAMUR (2005) and VDI/VDE (2006), a 'sensor' is a 'device that picks up physical measurement variables and converts them to standardized output signals'.

In VIM (2012), the term 'sensor' is defined as an 'element of a measuring system that is directly affected by a phenomenon, body, or substance carrying a quantity to be measured'. Some elementary examples give an explanation of this term: 'sensing coil of a platinum resistance thermometer, rotor of a turbine flow meter, Bourdon tube of a pressure gauge, etc'. In clause 3.8 of (VIM, 2012), it is stated that in some fields the term 'sensor' is replaced with the term 'detector', which, according to clause 3.9, has another meaning. The ambiguity of the term 'sensor' becomes stronger due to a vague definition of the 'measuring system' concept.

In essence, in all the documents listed, as well as in numerous dictionaries (which can be sourced with the help of the website: Anon., Sensor, http://universal_ru_en.academic.ru/2498046/), the term 'sensor' can be interpreted as an isolated device or a part of a more complicated device.

In practice, the term 'sensor' is applied for designating:

1. one sensitive element;
2. a primary measuring transducer that can include a sensitive element or a group of sensitive elements (e.g., an array);
3. a measuring transducer that consists of a number of separate transducers, connected in-series, for example, a primary transducer and amplifier;
4. an isolated unit that, in its case, can any component according to items (1)–(3) or a group of the corresponding components in any combination;
5. a unit, according to item (4), which, in its case, also contains additional signal processing units, for example, analogue-to-digital converter, bus interface, microcontroller and indicator (in any combination) (Taymanov and Sapozhnikova, 2009b).

Consequently, the term 'sensor' is vague. Therefore, in some publications the explanation of specific terms is given. For example, Duta and Henry (2005), state that 'for the aim of this paper a sensor is the device consisting of one or more transducers and a transmitter that converts the transducer signals into a form recognizable by the control or monitoring system' (p. 173).

The ambiguity of this term is aggravated due to development of sensor arrays, such as an artificial skin, electronic tongue, artificial nose and so on. In some publications (Abbate *et al.*, 2012; Gundlach, 2007; Spencer, 2012 and many others), in order to define a structurally complex device containing at least one sensor and additional transducers, the term 'sensor device' is applied. It is necessary to distinguish the concepts applied for a sensitive element (sensor) and a self-contained device including one or several sensors and other components.

1.2.2 Definitions of key terms related to devices with artificial intelligence

In scientific and technical publications, there is a great variety of terms used for devices with artificial intelligence (Taymanov and Sapozhnikova, 2009a, 2010a), for example, 'intelligent', 'smart', 'adaptive', 'self-validating', 'system-generating', 'self-calibrating', 'self-diagnosing', 'fault-tolerant', 'self-checking', 'cogent', 'brilliant' and so on (Allgood and Manges, 2001; Barberree, 2003; Bialas, 2010; Duta and Henry, 2005; Falconi *et al.*, 2007; Gaura and Newman, 2004; Hans and Ricken, 2007; Hashemian, 2005; Henry and Clarke, 1993; Huijsing, 2008; Huijising *et al.*, 1994; Hunter *et al.*, 2010; Itskovich, 2002; Jackson *et al.*, 2004; Johnson, http://www.telemonitor.com/doc/SmLegManu.PDF; Kirianaki *et al.*, 2002; Romanov, 1994; Wang *et al.*, 2005; Werthschutzky and Muller, 2007; Yurish, 2008, 2010; Zook *et al.*, 2006; Zvetkov, 1999; and many others). However, none of them was included in the International Vocabulary of Metrology (VIM, 2012).

A number of definitions of devices with artificial intelligence were stated in certain national standards, recommendations and international documents which do not refer to the sphere of metrology, (e.g., BSI, 2005; IEEE, 1997, 1999, 2003; NAMUR, 2005; VDI/VDE, 2006). In particular, in accordance with the IEEE Standard (1997, 1999, 2003), a 'smart sensor' is a 'sensor version of a smart transducer'. 'Smart transducer' is a 'transducer that provides functions over and above that necessary for generating a correct representation of a sensed or controlled physical quantity. This functionality typically simplifies the integration of the transducer into applications in a networked environment' (IEEE, 1997, 1999, 2003). It is unclear from this definition what additional functions are implied. Experience shows that the additional ('redundant') functions in question can be diverse. The importance of these functions is not the same for each customer, and the number of 'redundant' functions in a given device can vary. Therefore, at present such a definition of the term 'smart sensor' cannot satisfy customers.

In a number of publications, operations fulfilled in the past by a human operator are considered to be among the 'redundant' functions of such a smart sensor (e.g., Itskovich, 2002), in particular, zero adjustment, linearization, choosing a necessary measurement range, correcting for an influencing factor using a special sensor and so on. According to Romanov (1994), such devices should be considered as 'adaptive'. Specific features of 'sensors' and 'sensor systems' with self-adaptation capabilities are discussed in Yurish (2008, 2011).

According to Allgood and Manges (2001), a 'smart sensor' is a programmed 'measurement system that has sufficient computational capacity to support the data acquisition, memory and decision-making necessary to respond to algorithmic instructions'. It responds to calibration requests, reacts 'to interrogations about its health and status, and provides uncertainty estimations'. In Zook *et al.* (2006), the term 'smart sensor' implies that 'some degree of signal conditioning is included in the same package as the sensor. On the more sophisticated end of the spectrum, the 'sensor' unit can include devices with elaborated signal processing, displays and diagnostic or self-calibration features'.

According to Kirianaki (2002), a 'smart sensor is one chip, without external components, including the sensing, interfacing, signal processing and intelligence (self-testing, self-identification, self-validation or self-adaptation) functions'. However, this definition does not distinguish instruments with significantly different consumer features and complexity. For example, it levels:

- a device with self-check of a processing unit (secondary transducer) only and a device with self-check of both a primary transducer (sensor) and secondary transducer (transmitter);
- devices with self-identification and with self-adaptation.

According to Duta and Henry (2005), an 'intelligent sensor' can 'carry out internal diagnostics'. This enables a self-validating device to be developed that produces an estimation of measurement value, measurement uncertainty and 'a discrete valued flag indicating how the measurement value and uncertainty have been calculated'. In essence, a similar definition was used in Barberree (2003), where such a device is known as 'self-diagnosing', 'self-validating' or 'self-calibrating'. Wang *et al.* (2005) consider the intelligent instruments which are based on fieldbus and can perform such functions as self-calibration, self-compensation, self-validation and so on.

Hunter *et al.* (2010) discuss features of a 'smart sensor system' that may include multiple sensors and note the intelligent functions which can be performed 'at the sensor level including, but not limited to: self-calibration, self-health assessment, self-healing and compensated measurements (auto zero, calibration, temperature, pressure, relative humidity correction)'. 'The ideal goal is to have a self contained smart sensor system that is cost-effective, reliable, self-monitoring, reconfigurable, and can operate indefinitely'.

An alternative is to define smart and/or intelligent sensors by a description of their structure. In Di Lecce and Calabrese (2012) this approach is called 'hardware-centered'.

In some publications (e.g., Huijsing, 2008), a 'smart sensor' is considered to be a combination of 'a sensor, an analog interface circuit, an ADC, and a bus interface in one housing'. In the same publication, Huijsing also considers an 'integrated smart sensor'. It should contain 'one or more sensors, amplifiers, a chopper and multiplexers, an AD converter, buffers, a bus interface, addresses, and control and power management' in one chip. Most authors deem as 'intelligent' or 'smart' those devices which are equipped with a microprocessor or microcontroller (Anon. http://www.answers.com/topic/smart-sensor; Anon. http://computer.yourdictionary.com/smart-sensor; Bialas, 2010; Falconi, 2007; Gervais-Ducouret, 2012; Itskovich, 2002; Jackson, 2004; and many others). Such devices, as a rule, can carry out signal conditioning, filtering of noise, compensating for certain error components, producing an alert signal if a measured quantity value reaches a critical limit and so on. A parallel idea is presented in BSI (2005), where an intelligent instrument is considered to be a 'device where embedded computing capability is included as part of a measurement, control or actuation device'.

According to Gervais-Ducouret (2011), as opposed to any 'smart' sensor, an 'intelligent' sensor has a small microcontroller, a memory (flash, RAM, ROM) and optimized architecture for sensor applications. The intelligent sensor concept includes combination of possibilities to process sensor data, to reconfigure embedded functions and to aggregate external sensor data. The device with a microcontroller to which additional external sensors are connected for aggregating and – in a number of cases – joint pre-processing the data is called an 'intelligent sensor hub' (Di Lecce and Calabrese, 2012; Gervais-Ducouret, 2012). The intelligent sensor hub enables the processing of the data obtained from both an intrinsic sensor and a number of external (self-contained) sensors. Usually, it includes an extra interface (digital or analog).

Yurish (2010, 2011) emphasizes that the attribute 'smart', to a greater extent, can be related to 'technological' aspects of devices, while 'intelligent' is more applicable to 'functional' ones.

In industry, the most widespread attributes for a device with artificial intelligence are 'smart' and 'intelligent'. Meyer *et al.* (2009) and Kiritsis (2011), considering intelligent products, emphasize that the terms 'intelligent' and 'smart' can be applied interchangeably. An analysis of the 'intelligent' and 'smart' concepts in a number of widely used English, American and Russian dictionaries, testifies that sometimes both terms are applied in the same sense but, generally, the word 'intelligent' relates to a higher level of development, than 'smart'.

In 2009, a mini survey 'What does it mean 'smart sensor'?' was set up at the Sensors Web Portal. They received feedback from 227 participants from both academia and industry (Yurish, 2010, 2011). The survey results have shown that 18% of the participants named as 'smart' those 'sensors' which include combination of a sensing element, an analog interface circuit, ADC and bus interface. At least, 67% of experts considered the main features of 'smart sensors' to be some combination of self-identification, self-validation, self-testing, self-adaptation and/or self-calibration. Some participants emphasized the extended network capabilities of a 'smart sensor' (Yurish, 2010, 2011). It should be noted that the survey dealt only with the term 'smart', while the term 'intelligent' was not touched on. The responses of the survey participants do not indicate any differences in the interpretation of these terms, since the respondents had to consider them as synonyms.

Even so, the survey results do reveal the main features characterizing the trends in 'sensor' development. However, these results also demonstrate that there was no consensus between the experts with regard to the corresponding term (at least, this was so in 2009).

There is also a variety of terms and definitions related to MEMS (Enoksson *et al.*, 2005; Gaura and Newman, 2004; Madhavi, 2011; Newman *et al.*, 2006; Shkel, 2001; and others). In industrial applications, too, the most widely used attributes for the devices with artificial intelligence are 'smart' and 'intelligent'.

The ambiguity regarding terminology often causes a considerable confusion when customers try to make the best choice. Moreover, this ambiguity can become an obstacle to developing new equipment. Terms are blurred, acquiring multiple interpretations: when they were adopted, there were no ideas concerning their further usage. The same reason can offer an explanation of the polysemy and inconsistency of a nascent terminology. Therefore, the most important matter when designing a concept of an international vocabulary of terms is to take into consideration trends in developing measurement techniques and to reveal the problems which will determine the vocabulary of specialists in the field of measurements and control for the foreseeable future.

1.3 Key trends in the development of sensors (sensor devices) and MEMS

What is the driving force of sensor and MEMS development? Why are they called 'smart' or 'intelligent' with increasing frequency?

1.3.1 The method of analogy

To answer these questions, let us apply the analogy between biological organisms and technical means, as well as that of natures of their evolution. The methodical tool we present here seems to be of use, because the technical means are at a significantly earlier stage in their development than the most perfect biological organisms.

In their books, Wiener (1965) and Lem (1980) – who was not only a writer, but also an outstanding philosopher – demonstrated the efficiency of applying this analogy (Taymanov and Sapozhnikova, 2003, 2009a). At present, it is used with considerable success in evolutionary cybernetics (Red'ko, 2007; Turchin, 1977), as well as in bionics (biomimetics), which allows functions, structural elements and other features resembling similar solutions already found in nature to be realized by technical means (devices) (Bogue, 2009; Stroble *et al.*, 2009; Stroble Nagel *et al.*, 2010).

The similarity between biological and technical evolution forms not only a 'reference book' containing successful structural and functional decisions, it also gives a necessary strategic criterion for development in engineering (Taymanov and Sapozhnikova, 2008). Popper (1984) observed that all organisms reveal the answer to any problems – the problems having arrived with the beginning of life on Earth. Evolution presents the best solution.

1.3.2 Complication of organisms and sensors as a tendency of evolution

The key purpose of each living organism is to prolong its own life and the existence of its species. As life spans increase and geographic range extends, the number of possible dangerous situations increases, as does their variety. Accordingly, during the process of evolution the available methods with which to react to dangerous situations were also enriched. In nature, 'evolutionary change is not a continuous thing; rather it occurs in fits and starts, and it is not progressive or directional' (Ward, 2001). For example, organisms adapted, decreasing or increasing in size; the 'materials' forming a capsule, shell or coat changed. On the other hand, evolution 'has indeed shown at least one vector: toward increasing complexity' (Ward, 2001). First, came the conservative methods; later, adaptive methods arose; and, finally, methods based on intelligence came into being.

This vector is also traced in the development of biological sensors from the most basic (which perceive touch), to sensor arrays (moving from a sense of touch to a wide spectrum of perception) and then to multichannel sensor systems (which allow the estimation of more complicated multiparametric quantities – for example, color, odor and so on – including those distributed in space).

From an evolutionary point of view, there is no difference between animals and humans when considering methods for ensuring survival. 'Perhaps that is the fate of future human evolution: greater complexity through some combination of anatomy, physiology or behavior' (Ward, 2001). Such a point of view corresponds to the position shared by Asimov (1994), Hodos (1982), Klix (1980) and McFarland (1999), as well as many other biologists, philosophers and psychologists. According to this position, intelligence is a means of providing the most highly developed abilities in an organism's survival, including adaptation to current and future events (adaptation to climatic changes and intrinsic changes caused by ageing and so on). Therefore, intelligence enables the extension of life span – even over last several hundred years, human life span has doubled. It is logical to connect such extension with the increase in the role of intelligence in a human life, as well as with knowledge obtained using such intelligence.

In the process of evolution, it appears that intelligence has been the most powerful factor contributing to the survival of living creatures existing in a changing environment.

The similarity between biological and technical evolution is confirmed even in the finest details. The block-module principle of organization and development of molecular genetic systems shows that new systems are formed with blocks (modules) from bottom to top, macromolecular components playing the role of such modules (Red'ko, 2007).

Similar processes took place in the development of sensor devices with the capabilities to:

- interact with other devices,
- adapt to a changing environment,
- perform a self-check (in particular, metrological self-check for measuring instruments) and forecast sensor device health on the basis of artificial intelligence.

In sensor devices, separate components – such as sensitive elements, additional components, ADC, microcontrollers and so on – are combined to form a single entity.

By developing the idea of the analogy between the biological and technical evolution, it is possible to consider a direct analogy between the life span of living organisms and the functional lifetime of sensors and MEMS.

Over a long period of operation and in a wider field of application, various potential defects that were not detected during manufacture and also deviations in device characteristics may reveal themselves. Such defects can, for example, be caused by wear or by unforeseen changes in external factors. In sensors and MEMS used for the purpose of performing measurements, such defects and deviations in parameters lead to a growing lack of confidence in the reliability of measurement results. Accordingly, the definition 'intelligent' with respect to sensors and MEMS should be associated with the ability of these instruments to take active steps to counter unexpected influences.

Consequently, artificial intelligence at a certain level should contribute to widening the range of operational conditions and extending the lifetime of sensors and MEMS without any requirement for maintenance. For devices applied as measuring instruments, 'intelligence' should enable the reliability of measurements and calibration interval to be increased.

It is significant that the idea of applying artificial intelligence in sensors and MEMS (which are, as a rule, the weakest points of equipment) appeared and began to develop relatively recently. At this stage of technical evolution, it became possible to increase their structural complexity and it became necessary to extend their lifetime to a considerable degree (Taymanov and Sapozhnikova, 2010b).

Table 1.1 shows methods for providing the reliability of performance in the process of biological evolution (as applied to fauna) and technical evolution (as applied to sensors).

Certainly, new materials and technologies will be developed in the future. They will allow more stable characteristics to be realized. Undoubtedly, the requirements for reliability of sensors and MEMS will increase, too. Both these factors will result in keeping the developmental trends of their structures (from more simple structures to the more complex). It should be emphasized that, in general, the concept of 'complexity' has to take into account both structural and technological aspects. For example, to obtain a particular effect in a sensor, the application of a film with a spatially homogeneous structure (virtually, without any defects), or a component fabricated from extra-pure metal, should be considered as a sensor complication.

1.3.3 Features and forms of intelligence

As pointed out, intelligence is an efficient means for prolonging the existence of an intelligence 'carrier'. Intelligence implies an ability to perceive, analyze, accumulate and transmit information. The intelligence level depends on the extent of the completeness with which these features are realized.

Therefore, the first feature of intelligence is the capability of its 'carrier' to interact with the surrounding environment on the basis of the results of processed input information. For sensors and MEMS with artificial intelligence, applied in modern measuring and control systems, this means the capability to receive and transmit information under changing environmental conditions (in particular, in the presence of noise). It can be carried out on the basis of processing this information according to a given algorithm and transforming it into the relevant specific format.

The second feature is the ability of the intelligence 'carrier' to change its characteristics under the influence of information obtained, first, from the 'carrier's' sensing organs. For the devices with artificial intelligence under consideration, this feature requires the opportunity to change the param-

Table 1.1 Methods for providing functional reliability

Environment	Method	Examples	
		Biological evolution (as applied to fauna)	Technical evolution (as applied to sensors)
Stable	Conservative method for protection	Formation of capsules, shells, coats, etc.	Placing in a reliable housing
Slowly changing	Method of adaptation to a changing environment	Seasonal change of heat-protecting properties of a skin, slowing of life processes depending on the time of day or year	Active temperature stabilization, stabilization of a signal level due to introducing a negative feedback
	Method of interaction on the basis of data exchange	Informational interaction of living creature organs	Data exchange between sensors and other devices, as well as between sensors, including that with the help of sensor hubs
Rapidly changing	Method of adaptation to a changing environment	Change of a blood flow rate and pulse rate depending on a load	Temperature correction, automatic selection of a required measurement range
	Method, based on intelligence	Development of sense organs, health check, forecast of vitally important situations, ensuring survival	Increase in the number and variety of sensors in a sensor of device, as well as external sensors connected to it for joint data processing; self-check and forecasting the health of the sensor device, including metrological health; self-correction

eters and/or the operation algorithm in accordance with signals obtained from embedded sensors or external sources.

The third feature is a capability of the intelligence 'carrier' to estimate its own health. For sensors and MEMS with artificial intelligence, this feature corresponds to a self-check ability, including a metrological self-check ability, if such devices are applied as measuring instruments. In many cases, the results of the metrological self-check can be used to perform self-correction and self-recovery, as well as obtaining a forecast of metrological health.

It is noteworthy that the three features mentioned can be related both to humans and to devices with artificial intelligence. Accumulation of such features demonstrates that the level of natural or artificial intelligence of its 'carrier' increases. The features observed are reminiscent, to some extent, of levels of product intelligence according to Kiritsis (2011), but Kiritsis does not consider the analogy of products and living creatures.

In the hierarchy of these features, the third is the most important. In essence, it is a question of an ability to check the reliability of information obtained after it has been processed, which is the main characteristic of a 'carrier' of intelligence.

In nature, intelligence has developed itself in two ways: formation of a 'collective mind' of many living organisms, and the development of intelligence of a separate individual. If there is a high risk of extinction for individual living organisms, the 'collective mind' provides a way of preserving the experience gained and supporting the life of the species as a whole. Certainly, this is the type of intelligence which originated in the early stages of evolution, when the life span of an individual organism was short. A typical example is given by swarming insects; e.g., bees, which select reliable information by 'voting'. The importance and validity of information obtained by scout bees depends on the number of bees that obtained the information (McFarland, 1999). Incidentally, origination of a 'democratic method' in human society for estimating the reliability of vitally important information and the subsequent development of this method for controlling a state demonstrate the efficiency of applying experience from nature (Taymanov and Sapozhnikova, 2008).

Formation of the 'collective mind' is a typical example of a meta-system conversion representing the integration of a number of autonomous subsystems at a lower level (they can, to some extent, differ) and the development of an additional control mechanism at a higher level (Red'ko, 2007). A similar approach is frequently used in metrological practice for checking the reliability of information. For example, in nuclear power plants, sensors, the quantity of which is redundant, are combined into a multichannel system (Hashemian, 2006). The metrological serviceability of a sensor can be estimated on the basis of deviation of its signal from the signals of the remaining ones included in the system. Information from the majority of the sensors is considered to be valid. The 'swarm' model is applied (e.g., in geology and other fields) to reveal reliable information (Bogue, 2009).

However, it has transpired that the selection of reliable information on the basis of the 'swarm' model is not the most effective choice in biological evolution due to insufficient flexibility in a changing environment (Taymanov and Sapozhnikova, 2008).

The development of intelligence of an individual living creature was accompanied by an increase in its life span. Intelligence has been formed within the framework of a meta-system development as a result of the integration of a number of independent nervous structures, each of them linked with one or several biological sensors and distinguishing themselves by individual features. The 'health' of these sensors is checked with the help of intelligence. In comparison with the 'collective mind', the intelligence of an individual has an advantage in searching for effective ways to survive in a changing environment.

It is considered that, in addition to the minimum of 'sensors' for the quantity to be measured, a sense organ also has supplementary sensors. The brain has a special mechanism for testing the stability of essential activity characteristics. This mechanism, known as an 'error detector', was discovered by the famous Russian Academician Bechtereva (Bechtereva *et al.*, 2005). Initially a person diagnoses a 'malfunction' of his organism – including a 'malfunction' of a sense organ, such as an eye or ear – either through an unpleasant sensation caused by signals coming from these supplementary sensors, or indirectly from the 'error detector'. Following this, an appropriate decision is made that will resolve the issue (e.g., to irrigate an eye or take medicine). Substantially, diagnosing an indisposition and making the decision is an interaction of the brain with individual sense organs. In essence, this interaction does not contradict the classical definition of the intellectuality of a technical device as being an ability based on knowledge to communicate with a customer and make a decision with respect to a mode of future actions (Andrew, 1983; Romanov, 1994).

It should be noted that to select external information of vital importance (e.g., video- or audio-information) these supplementary signals from the sense organs are not necessary and, in this sense, they are redundant.

Besides the method discussed earlier, a human, together with other developed living creatures, uses auxiliary means to detect deterioration in the operation of an individual's sense organs:

- analysis of total information, obtained through a group of various organs of sense;
- response of other members of a society.

In other words, living creatures with developed intelligence use both a 'swarm of sense organs' and a 'swarm of individuals' to check 'measurement information' reliability. However, these 'swarm' opportunities are only an auxiliary means.

Thus, in biological evolution, both types of intelligence considered in this chapter develop in parallel. Sometimes, they supplement each other, but the intelligence of individual has gained priority and improved at a greater pace. In human society, the 'collective mind' is a stabilizing factor; it plays the same role in multichannel systems with artificial intelligence. The intelligence of individual provides a greater ability to survive both for a particular individual organism and for the species as a whole.

The similarities between biological and technical evolution provide the basis for forecasting future trends in sensor devices and MEMS. For example, in the short term, one should expect the development of self-studying intelligent devices with the ability to self-check their health. Such instruments will be able not only to correct consequences of 'ageing' and influencing factors, but also to adjust the values of their parameters automatically on the basis of predicted changes. This can further enhance the efficiency of their operation. In technical complexes, sensor devices with the ability to self-check their health, combined into a system with a 'collective mind', will be used widely. Soon, arrays that are sensitive to one quantity only, as well as arrays combining groups of sensors susceptible to a variety of quantities, will find a wider application (Taymanov and Sapozhnikova, 2008, 2009b).

1.4 Suggestions for improving terminology in the field

The concept of the history and trends of sensor and MEMS development, based on the analogy between biological and technical evolution, offers a firm foundation for developing a systematized terminology to overcome or, at least, to reduce the existing ambiguity.

Below, we present a proposed group of the most important terms (in bold) and their definitions related to the subject of the present book, together with brief comments.

Sensor: an element of a measuring or control instrument, which is directly affected by a phenomenon, body, or substance carrying a quantity to be measured. It is the simplest sensitive element. (The definition is close to the definition given in the International Vocabulary of Metrology (VIM, 2012) and corresponds to item (a) of the list given in Section 1.2.1).

Sensor device: a self-contained device including one or more sensors, and in some cases additional components (e.g., an amplifier, filter, ADC, interface, and so on). This term is close to the Russian term *datchik*.

Measuring system: a group of sensor devices and other components associated with them, having a common data processing unit, for measuring one or several quantities distributed in space (space can be one-, two- or three-dimensional).

Metrological self-check of a sensor device: an automatic procedure for testing whether the uncertainty or error of a sensor device in the course of its operation lies within permissible limits specified under operating conditions. This procedure is carried out using a reference value generated with the help of an additional (redundant) embedded component (a sensor, secondary transducer (transmitter) or material measure) or an additional parameter for an output signal.

The term 'reference value' corresponds to the same term given in VIM (2012). The metrological self-check can be realized in two forms: a direct or diagnostic self-check (Taymanov and Sapozhnikova, 2005; Taymanov *et al*., 2011). If the metrological self-check is accompanied by evaluation of uncertainty or error, it is usually called 'self-validation' (Sapozhnikova *et al*., 2005).

Smart sensor device: a sensor device capable of interaction with other devices on the basis of processing the data received. Such a definition corresponds to the first (simplest) feature of intelligence given in Section 1.3.3 of this chapter.

Here, the concept 'interaction' suggests the possibility of noise-immune communication with other remote devices. In turn, noise-immune communication, as a rule, is realized by coding information (an analog signal is converted into a digital form in the sensor device). This provides reliability of information when transmitting it over longer distances. By these means, the 'survivability' of a sensor device in a changing environment can be achieved.

An example would be a sensor device containing a sensor, ADC, interface for communication with a common bus, and certain other components.

Smart sensor hub: a 'smart sensor device' to which additional external sensors are connected for joint data processing. This instrument may be considered as a measuring system.

Adaptive sensor device: a sensor device, the parameters and/or operative algorithms of which can change in the course of operation, depending on internal signals from additional sensors and secondary transducers (transmitters) this device contains.

An adaptive sensor device is capable of response (adjustment) to a measurement range, speed of measurand variation, effect of influencing quantities and so on, within the limits specified in technical documentation. The definition of an 'adaptive' sensor device supposes realization of the second feature of intelligence.

All the capabilities listed indicate the working capacity of an adaptive sensor device within an expanded range of environmental conditions.

Adaptation to changing conditions makes it possible to correct output signals according to a pre-set interaction. Thus, this adaptation contributes an extension of the maintenance-free working life of sensor devices and MEMS. The reliability of measurement results obtained with adaptive sensor devices applied as measuring instruments is increased, and the calibration interval can be extended.

In a sensor device, two intelligence features can be combined; that is, the sensor device can be simultaneously 'smart' and 'adaptive'.

Intelligent sensor device: a sensor device with a self-check function.

At present, this device is an instrument that enables the highest level of information reliability. According to the above classification, it can be characterized by all three features.

In the intelligent sensor device applied as a measuring instrument (Taymanov *et al.*, 2011), it is possible to do the following:

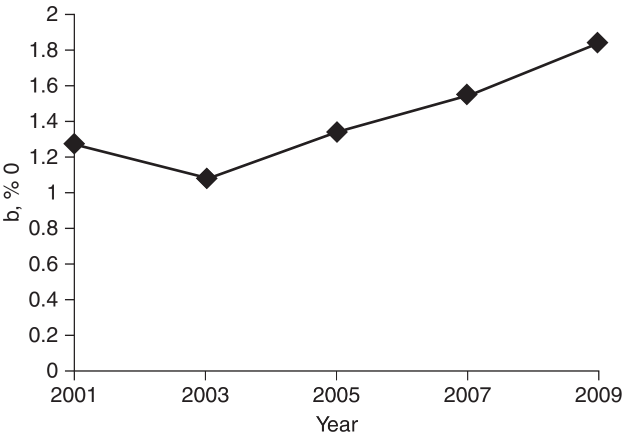
- to realize the metrological self-checking;
- to forecast the metrological health of the sensor device, taking into account the results of self-checking;
- to correct automatically the uncertainty or error caused by influencing quantities, the value of which is unknown, and/or ageing of certain device components (in a number of cases);
- to extend the calibration interval on the basis of the results of self-checking, forecasting metrological health and automatic correction;
- to perform the self-recovery of the sensor device (This can be done in some cases when this device has a single defect).

Self-recovery: an automatic procedure for damping the consequences in the event that a defect has arisen in a sensor device; that is, it is a procedure of providing a necessary level of fault-tolerance.

Fault-tolerance: the ability of a sensor device to keep specified characteristics within permissible limits in the event that a single defect has occurred in it.

Methods for metrological self-checking and examples of intelligent sensor devices are considered, in particular, in Taymanov *et al.* (2011).

Intelligent sensor devices provide the greatest maintenance-free lifetime and optimum efficiency within a wider range of environmental conditions. Unfortunately, the semantic difference between the terms 'intelligent' and 'smart', when they are translated into a number of wide spread languages (e.g., German and Russian) is significantly less than the difference in performance characteristics of the corresponding sensor devices. This can lead to certain difficulties.



1.1 Variation of interest the specialists show in the development of intelligent sensor devices, as a percentage of the total interest in measurement or control.

Sometimes it is necessary to locate a sensor device separate and apart from a data processing unit. This can be because of a harsh operating environment, which is unacceptable for the microprocessors in a sensor device. The same situation can also arise if centralized data acquisition is required in a measuring system. In this case, a data-redundant or adaptable sensor device can be developed:

Data-redundant sensor device: a sensor device enabling a reference value to be obtained on the basis of an additional output signal parameter or of an additional built-in component. In the event of connecting such a sensor device to an external signal processing unit, the metrological self-check can be realized.

Adaptable sensor device: a sensor device, parameters and/or operation algorithm of which can be changed by external signals in the course of the operation process. In the event that a sensor device is connected to an external signal processing unit, it can operate as an adaptive sensor device.

The main features of the definitions given above can be applied to a measuring system, particular forms of which can be 'smart', 'adaptive' and 'intelligent' depending on sensor devices and MEMS which are used in this system.

Corresponding proposals with regard to the terms applied in metrology were prepared for a future edition of the International Vocabulary of

Metrology. Some of the terms and definitions given above were included in the Russian standard (GOST R 8.673, 2009).

Industrial production of sensors and MEMS with artificial intelligence does not keep pace of new developments in this field. Various smart sensor have been produced for many years in large batches. As for adaptive sensor devices, their range is significantly less. At present, intelligent sensor devices are produced in small individual batches. Figure 1.1 illustrates the increase in interest shown by specialists, particularly in the development of intelligent sensor devices corresponding to the definition given in this chapter. The statistical data were obtained using the 'Scopus' (a search engine for researching scientific information on the Internet). The vertical axis presents the percentage of publications in the field of measurement or control, that contain the term 'metrological self-check' or words close to it in meaning.

1.5 Conclusion

Sensors and MEMS with artificial intelligence permit operational costs to be reduced but, in so doing, cause a number of new problems. To resolve these problems, it is necessary to provide specialists with the opportunity to share their experience, to define the performance requirements and characteristics of new devices, and to develop new normative documents (standards, guides and recommendations).

The analysis of the terms, considered in this chapter, which are applied to sensors and MEMS with artificial intelligence, shows that an ambiguity in terminology impedes the resolution of these tasks. In this regard, harmonization of the terminology is of significant importance.

Arguments for applying the analogy between biological and technical evolution have been given in order to reveal the trends in sensor development. A group of interrelated terms and their definitions has been proposed on the basis of this analogy, taking into consideration the results of terminological analysis.

1.6 Acknowledgment

The authors of the chapter express their deep gratitude to Dr Nihtianov for useful and interesting discussions.

1.7 References

Abbate, S., Avvenuti, M. and Light, J. (2012), 'MIMS: a minimally invasive monitoring sensor platform, *IEEE Sensors Journal*, Vol. **12**, No 3, art. no. 5762310, 677–684.

Allgood, G. O. and Manges, W. W. (2001), 'Sensor Agents – When Engineering Emulates Human Behavior', *Sensors*, August, 2001. Available from: http://

www.sensorsmag.com/sensors-mag/sensor-agents-when-engineering-emulates-human-behavior-781 (Accessed 27 August 2013).

Andrew, A. M. (1983), *Artificial Intelligence* (Cybernetics and systems series), Tunbridge Wells, Kent, Abacus Press.

Anon. 'Sensor,' Available from: Anon., Sensor, http://www.onelook.com/?w=sensor&ls=a (Accessed 27 August 2013).

Anon. 'Smart Sensor', Available from: http://www.answers.com/topic/smart-sensor (Accessed 27 August 2013).

Anon. 'Smart sensor – technical definition'. Available from: http://computer.yourdictionary.com/smart-sensor (Accessed 27 August 2013).

Asimov, I. (1994), *The Human Brain. Its Capacities and Functions*, Signet (MM), Revised Expanded edn, Boston, Massachusetts, Houghton Mifflin Company.

Barberree, D. (2003), 'Dynamically self–validating contact temperature sensors', in *Proceedings of the Conference 'Temperature: Its Measurement and Control in Science and Industry'*, 21–24 October 2002, Chicago, USA. Vol. 7, Melville, New York, AIP Conference Proceedings, 1097–1102.

Bechtereva, N. P., Shemyakina, N. V., Starchenko, M. G., Danko, S. G. and Medvedev, S. V. (2005), 'Error detection mechanisms of the brain: background and prospects', *International Journal of Psychophysiology*, Vol. **58**, No. 2–3, 227–234.

Bialas, A. (2010), 'Common criteria related security design patterns – validation on the intelligent sensor example designed for mine environment', *Sensors*, Vol. **10**, No. 5, 4456–4496.

Bogue, R. (2009), 'Inspired by nature: developments in biomimetic sensors', *Sensor Review*, Vol. **29**, No. 2, 107–111.

BSI (2005), *Specification for Data Quality Metrics of Industrial Measurement and Control Systems*, BS7986: 2005, British Standards Institute, 389 Chiswick High Road, London W4 4AL.

Di Lecce, V. and Calabrese, M. (2012), 'From smart to intelligent sensors: a case study', *Sensors & Transducers Journal*, Vol. **14-1**, No. 3, special issue, 1–17, http://www.sensorsportal.com/HTML/DIGEST/P_SI_178.htm (Accessed 27 August 2013).

Duta, M and Henry, M. (2005), 'The fusion of redundant SEVA measurements', *IEEE Transactions on Control Systems Technology*, Vol. **13**, No. 2, 173–184.

Enoksson, P., Rusu, C., Sanz–Velasco, A., Bring, M., Nafari, A. and Bengtsson, S. (2005), 'Wafer bonding for MEMS', in *Proceedings – Electrochemical Society*, PV 2005–02, 157–172.

Falconi, C., Martinelli, E., Di Natale, C., D'Amico, A., Maloberti, F., Malcovati, P., Baschirotto, A., Stornelli, V. and Ferri, G. (2007), 'Electronic interfaces', *Sensors and Actuators, B: Chemical*, Vol. **121**, No. 1, 295–329.

Gaura, E. and Newman, R. M. (2004), 'Smart, intelligent and cogent microsensors – intelligence for sensors and sensors for intelligence', in *Proceedings NSTI Nanotechnology Conference and Trade Show (NSTI Nanotech 2004)*, March 7–11, 2004, Boston, Massachusetts, USA, 1, 443–446.

Gervais-Ducouret, S. (2011), 'Next smart sensors generation', in *Proceedings of the IEEE Sensors Applications Symposium (SAS' 11)*, 22–24 February 2011, San Antonio, Texas, USA, 193–196.

Gervais-Ducouret, S. (2012), 'Smart & intelligent sensors', Workshop on the Intelligent Sensor Hub, Sensordevices-WISH-2012, 19–24 August 2012, Rome, Italy.

GOST R 8.673 (2009), State System For Ensuring The Uniformity of Measurements. Intelligent Sensors and Intelligent Measuring Systems. Basic Terms and Definitions, Moscow, Standartinform, 2010.

Gundlach, D. J. (2007), 'Organic electronics: low power, high impact', *Nature Materials*, Vol. **6**, No. 3, 173–174.

Hans, V. and Ricken, O. (2007), 'Self-monitoring and self-calibrating gas flow meter', in *Proceedings of the 8th International Symposium on Measurement Technology and Intelligent Instruments*, 24–27 September 2007, Sendai, Japan, 285–288.

Hashemian, H. M. (2005), *Sensor Performance and Reliability*, USA, ISA.

Hashemian, H. M. (2006), *Maintenance of Process Instrumentation in Nuclear Power Plants*. Berlin, Heidelberg, New York: Springer.

Henry, M. P. and Clarke, D. W. (1993), 'The self–validating sensor: Rationale, definitions and examples', *Control Engineering Practice*, Vol. **1**, No. 4, 585–610.

Huijising, J. H., Riedijk, F. R. and van der Horn, G. (1994), 'Developments in integrated smart sensors', *Sensors and Actuators*, *A: Physical*, Vol. **43**, No. 1–3, 276–288.

Huijsing, J. H. (2008), 'Smart sensor systems: Why? Where? How?', in *Smart Sensor Systems*, G. C. M. Meijer (ed.), Chichester, UK, John Wiley & Sons.

Hodos, W. (1982), 'Some perspectives on the evolution of intelligence and the brain', in: *Animal Mind–Human Mind*, D. R. Griffin (ed.), Berlin, Springer Werlag.

Hunter, G. W., Stetter, J. R., Hesketh, P. J. and Liu, C. C. (2010), 'Smart sensor systems', *Electrochemical Society Interface,* Vol. **19**, No. 4, 29–34.

IEEE Standard for a Smart Transducer Interface for Sensors and Actuators – Digital Communication and Transducer Electronic Data Sheet (TEDS) Formats for Distributed Multidrop Systems (2003), std. 1451.3.

IEEE Standard for a Smart Transducer Interface for Sensors and Actuators – Network Capable Application Processor (NCAP) Information Model (1999), std. 1451.1.

IEEE Standard for a Smart Transducer Interface for Sensors and Actuators – Transducer to Microprocessor Communication Protocols and Transducer Electronic Data Sheet (TEDS) Formats (1997), std. 1451.2.

Itskovich, E. L. (2002), 'The modern intelligent industrial sensors', *Sensors and Systems*, No. 2, 42–47.

Jackson, R. G. (2004), *Novel Sensors and Sensing*, Bristol and Philadelphia, Institute of Physics Publishing.

Johnson, R. N. (2002), 'Applying smart sensor technology to existing real-world (legacy) systems'. Available from: http://www.telemonitor.com/doc/SmLegManu.PDF (Accessed 27 August 2013).

Kiritsis, D. (2011), 'Closed–loop PLM for intelligent products in the era of the internet of things', *CAD Computer Aided Design,* Vol. **43**, No. 5, 479–501.

Kirianaki, N. V., Yurish, S. Y., Shpak, N. O. and Deynega, V. P. (2002), *Data Acquisition and Signal Processing for Smart Sensors*, Chichester, England, John Wiley & Sons.

Klix, F. (1980), *Erwachendes Denken. Eine Entwicklungsgeschichte Der Menschlichen Intelligenz*, VEB Deutscher Verlag der Wissenschaften.

Kuhn, T. S. (1962), *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press.

Lem, S. (1980), *Summa Technologiae*, Berlin, Verlag Volk & Welt.

Madhavi, K. Y., Sumithradevi, K. A., Krishna, M. and Dharani, A. (2011), 'Diaphragm design for MEMS pressure sensors using a data mining tool', in *Proceedings of the World Congress on Engineering 2011 (WCE 2011)*, 2, 1353–1356.

McFarland, D. (1999), *Animal Behaviour*: *Psycology, Ethology, and Evolution*, Hertfordshire, Prentice Hall.

Meyer, G. G., Framling, K. and Holmstrom, J. (2009), 'Intelligent products: a survey', *Computers in Industry 2009,* Vol. **60**, No. 3, 137–148.

NAMUR (2005), 'Self–monitoring and Diagnosis of Field Devices', *Recommendation NE 107,* Leverkusen.

Newman, R., Gaura, E. and Mount, S. (2006), 'The myth and reality of wireless sensor networks: Designing optimally redundant networks', in *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA 2006)*, art. no. 4026183, 780–786.

Popper, K. (1984), *An Evolutionary Epistemology. Evolutionary Theory: Paths into the Future*, in J. W. Pollard (ed.), Chichester and New York, John Wiley & Sons, 239–255.

Red'ko, V. G. (2007), *Evolution. Neural Networks. Intelligence. Models and Concepts of the Evolutionary Cybernetics*, Moscow, KomKniga.

Romanov, V. N., Sobolev, V. S. and Zvetkov, E. I. (1994), *Intelligent Measurement Instruments*, Moscow, Tatijanin den'.

Intergovernmental Recommendation RMG 29–99 (2000), State System for Ensuring The Uniformity of Measurements. Metrology. Basic Terms and Definitions, Minsk, Intergovernmental Committee on Standardization, Metrology, and Certification, 2000.

Sapozhnikova, K., Henry, M. and Taymanov, R. (2005) '*The Need for Standards in Self–diagnosing and Self–validating Instrumentation*', *Joint International IMEKO TC1+TC7 Symposium,* 21–24 September 2005, Ilmenau, Germany (CD–ROM).

Shkel, A. M. (2001), Smart MEMS: 'Micro-structures with error-suppression and self-calibration control capabilities', in *Proceedings of the American Control Conference*, 25–27 June 2001, Arlington, VA, USA, 2, 1208–1213.

Spencer, M. J. S. (2012), 'Gas sensing applications of 1D-nanostructured zinc oxide: insights from density functional theory calculations', *Progress in Materials Science*, Vol. **57**, No. 3, 437–486.

Stroble, J. K., Stone, R. B. and Watkins, S. E. (2009), 'An overview of biomimetic sensor technology', *Sensor Review*, Vol. **29**, No. 2, 112–119.

Stroble Nagel, J. K., Stone, R. B. and McAdams, D. A. (2010), 'Function-based biology inspired concept generation', in *Biomimetics Learning From Nature*, A. Mukherjee (ed.), InTech, 94–116. Available from: http://sciyo.com/books/show/title/biomimetics-learning-from-nature. ISBN 978-953-307-025-4, DOI: 10.5772/55321, (Accessed 29 August 2013).

Tarbeyev, Yu., Kuzin, A., Taymanov, R. and Lukashev, A. (2007), 'New stage in the metrological provision for sensors', *Measurement Techniques*, Vol. **50**, No. 3, 344–349.

Taymanov, R. and Sapozhnikova, K. (2003), 'Intelligent measuring instruments. maximum reliability of measuring information, minimum metrological maintenance', in *Proceedings of the XVII IMEKO World Congress*, Dubrovnik, Croatia, 22–27

June, 2003, CD 1 0623 20 3030C314, 1094–7. Available from: http://www.imeko.org/ (Accessed 29 August 2013).

Taymanov, R. and Sapozhnikova, K. (2005), 'Intellectualization of the built–in measuring instrument as the way to enhance the reliability of equipment', in *Problems of Mechanical Engineering: Precision, Friction, and Depreciation, Reliability, and New Technologies*, V, P. Bulatov (ed.), St. Petersburg, Nauka, 421–469.

Taymanov, R. and Sapozhnikova, K. (2008), 'Metrological self-check and evolution of metrology', in *Proceedings of the 12th IMEKO TC1 – TC 7 Joint Symposium*, 3–5 September, Annecy, France, 212–217. Available from: http://www.imeko.org/ (Accessed 29 August 2013).

Taymanov, R. and Sapozhnikova, K. (2009a) 'Problems of terminology in the field of measuring instruments with elements of artificial intelligence', *Sensors & Transducers Journal*, Vol. **102**, No. 3, 51–61. Available from: http://www.sensorsportal.com/HTML/DIGEST/P_401.htm (Accessed 29 August 2013).

Taymanov, R. and Sapozhnikova, K. (2009b), 'Problems of terminology improvement in metrology', in *Proceedings of the XIX IMEKO World Congress on Fundamental and Applied Metrology,* Lisbon, Portugal, 6–11 September 2009, 1080–1085. Available from: http://www.imeko.org/ (Accessed 29 August 2013).

Taymanov, R. and Sapozhnikova, K. (2010a) 'Problems of terminology in the field of intelligent sensors and systems', *Key Engineering Materials.* Switzerland, Trans Tech Publications, 437, 434–438.

Taymanov, R. and Sapozhnikova, K. (2010b) 'Metrological self–check and evolution of metrology', *Measurement*, Vol. **43**, No. 7, 869–877, DOI:10.1016/j.measurement.2010.04.004.

Taymanov, R., Sapozhnikova, K. and Druzhinin, I. (2011), 'Sensor devices with metrological self–check', *Sensors & Transducers Journal*, Vol. **10** (special issue), No. 2 (February 2011), 30–44. Available from: http://www.sensorsportal.com/HTML/DIGEST/P_SI_131.htm (Accessed 29 August 2013).

Turchin, V. F. (1977), *The Phenomenon of Science. A Cybernetic Approach to Human Evolution*, New York, Columbia University Press.

VIM. *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms* (2012), 3rd edn, 2008 version with minor corrections, BIPM, JCGM 200. Available from: http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf (Accessed 29 August 2013).

VDI/VDE Guideline 2650 (2006), Requirements for Self–monitoring and Diagnostics in Field Instrumentation. General requirements. Part 1, Dusseldorf, VDI/VDE.

Wang, B.-S., Zhang, J., Ma, Y.- G. and Cao, W. -L. (2005), 'Research of MBC based on fieldbus and intelligent instrument', in *Proceedings of SPIE – The International Society for Optical Engineering*, 6042, Y. Wei, K. T. Chong, T. Takahashi, S. Liu, Z. Li, Z. Jiang and J. Y. Choi (eds), art. no. 60421Z.

Ward, P. (2001), *Future Evolution*, New York, Nimes Books, Henry Holt & Company.

Werthschutzky, R. and Muller, R. (2007), 'Sensor self–monitoring and fault–tolerance', *Technisches Messen*, Vol. **74**, No. 4, 176–184.

Wiener, N. (1965), *Cybernetics: or Control and Communication in the Animal and the Machine*, 2nd edn, Cambridge, MA, MIT Press.

Yurish, S. Y. (2008), 'Self–adaptive smart sensors and sensor systems', *Sensors & Transducers Journal*, Vol. **94**, No. 7, 1–14.

Yurish, S. Y. (2010), 'Sensors: smart vs. intelligent', *Sensors & Transducers Journal*, Vol. **114**, No. 3, I–VI.

Yurish, S. Y. (2011), *Digital Sensors and Sensor Systems: Practical Design*, IFSA. Available at ifsa.books@sensorsportal.com.

Zook, D., Bonne, U. and Samad, T. (2006), 'Sensors in control systems', in *Control Systems, Robotics, and Automation*, I. H. Unbehauen (ed), *Encyclopedia of Life Support Systems (EOLSS)*, developed under the Auspices of UNESCO, Oxford, UK, EOLSS Publications.

Zvetkov, E. I. (1999), 'Intellectualization of measurement instruments', in *Proceedings of the International Conference 'Soft Calculations and Measurements SCM–99',* 25–28 May 1999, St. Petersburg, Russia, 42–46.

2

Direct interface circuits for sensors

F. REVERTER, Universitat Politècnica de Catalunya
(UPC), Spain

DOI: 10.1533/9780857099297.1.27

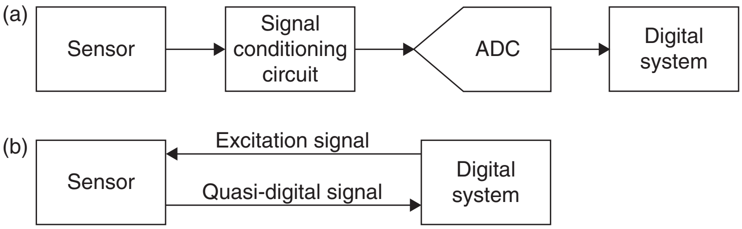
Abstract: This chapter discusses the direct connection of analogue sensors to digital systems without using any analogue circuit in the signal path. It first reviews the operating principle of analogue sensors (mainly, resistive and capacitive sensors) and digital systems (basically, microcontrollers). It then explains how to connect different sensor topologies to a microcontroller to build a direct interface circuit. Finally, it discusses some applications using commercial devices.

Key words: sensor electronic interface, microcontroller, resistive sensor, capacitive sensor.

2.1 Introduction

Just as human beings acquire information on their environment through their senses and process such information using their brain, electronic systems try to do the same by means of sensors and processing digital devices such as microcontrollers (μC) or microprocessors (μP). Nowadays, sensors and processing devices have become essential for the smooth running of our lives and they are present in many fields: industrial, automobiles, aircraft, medical devices, consumer electronics and home appliances, among others.

The classic block diagram of a sensor electronic interface is shown in Fig. 2.1a (Pallàs-Areny and Webster, 2001). First, information about the measurand (e.g., temperature) is converted to the electrical domain by means of the sensor, which usually provides an analogue electrical signal of low amplitude that carries some noise. Afterwards, the signal conditioning circuit, which generally relies on operational amplifiers, performs some or all of the following tasks in the analogue domain: sensor output-to-voltage conversion, amplification, filtering, linearization and/or demodulation. The resulting analogue signal is then digitized via an analogue-to-digital converter (ADC). Finally, a digital system (e.g., a μC) acquires, stores, processes, controls, communicates (to other devices or systems) and/or displays the digital value with information about the measurand.



2.1 (a) Classic block diagram of a sensor electronic interface, (b) direct interface circuit.

Thanks to the rapid advances in integrated circuit (IC) technologies, nowadays several blocks shown in Fig. 2.1a can be integrated into the same chip. There are commercially available ICs that integrate the processing digital system, the ADC and/or a signal conditioning circuit (for instance: MSC1210 from Texas Instruments). There are also commercial ICs that integrate the sensor, its signal conditioning circuit and/or the ADC (e.g., ADXL103 and ADXL312 accelerometers from Analog Devices); such chips are usually known as integrated smart sensors (Huising, 2008). Other commercial ICs integrate the required signal conditioning circuit and the ADC to measure a specific type of sensor (for instance: ADS1232 from Texas Instruments for bridge-type resistive sensors and AD7745 from Analog Devices for single and differential capacitive sensors).

For some sensors, however, the block diagram in Fig. 2.1a can be simplified to that shown in Fig. 2.1b (Reverter and Pallàs-Areny, 2005a), where the sensor is directly connected to the digital system without using either the signal conditioning circuit or the ADC; these circuits were initially proposed in application notes of μC's manufacturers (Bierl, 1996; Cox, 1997; Richey, 1997). In such a circuit topology (a so-called 'direct interface circuit'), the digital system appropriately excites the analogue sensor in order to obtain a signal that can be directly measured in the digital domain; for instance, using an embedded digital timer. In comparison with the sensor electronic interface shown in Fig. 2.1a, a direct interface circuit is simpler and needs fewer components; actually, it just needs a common low-cost general-purpose 8-bit μC working as a digital system. Therefore, a direct interface circuit has advantages in terms of cost, physical space and power consumption, which is of major interest in autonomous sensors and in wireless sensor networks. Furthermore, the performance of such circuits is quite remarkable taking into account their simplicity; for instance: a non-linearity error of 0.01% full-scale span (FSS) and a resolution of 13 bits when measuring resistive sensors in the kiloohm range, and 0.1% FSS and 9 bits when measuring capacitive sensors in the picofarad range. One of the main limitations of such circuits is that the measuring time depends on the value of the measurand and can be quite long (e.g., about units or tens of millisecond).

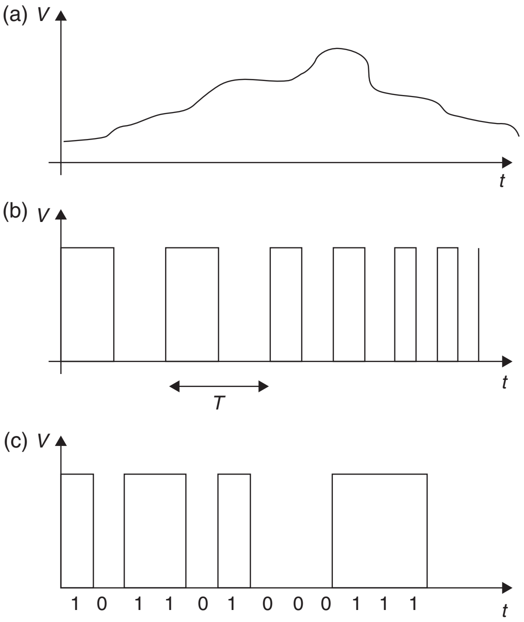
This chapter deals with the direct interface circuit topology shown in Fig. 2.1b and is organized as follows. Section 2.2 provides an overview of the sensors that can be measured using such circuits. Section 2.3 describes the needs of the digital system (especially, a μC) to perform the tasks required in such circuits. After describing sensors and μCs, Section 2.4 explains how to join them to build a direct interface circuit. Section 2.5 shows some applications of these circuits. Section 2.6 describes likely future trends on this topic. Finally, Section 2.7 provides additional information about research groups working on these circuits.

2.2 Sensors

Sensors are the first block of the measurement chain (see Fig. 2.1) and are in charge of providing a signal in the electrical domain with information about the measurand. Sensors can be classified in different ways. Common classifications are in terms of the energy domain of the measurand (e.g., thermal sensors) or in terms of the electrical signal provided at the output (e.g., resistive sensors). However, Kirianaki *et al.* (2002) have proposed to classify them in terms of the domain of the electrical output signal. Accordingly, we have:

- *Analogue sensors*, which provide an analogue output signal whose amplitude is modulated by the measurand (see Fig. 2.2a). Such sensors have been, and still are, very common in industrial applications (e.g., the standardized Pt100 temperature sensor).
- *Quasi-digital sensors*, which provide an analogue or digital output signal with a time-related parameter (such as period, frequency, duty cycle or time interval) that is modulated by the measurand (see Fig. 2.2b). Although such time-related information is analogue, it can be directly measured by a digital system with a timer or a counter, hence the name 'quasi digital'. Such output signals can be the result of the sensing principle (e.g., a resonant sensor) or obtained by embedding an analogue sensor and some additional circuitry (e.g., an oscillator) into the same IC sensor (e.g., TMP04 from Analog Devices).
- *Digital sensors*, which provide a digital output signal (see Fig. 2.2c) with a sequence of logic levels (i.e., 0s and 1s) modulated by the measurand. Such signals can also be the result of the sensing principle (e.g., an absolute position optical encoder) or obtained by means of an IC sensor that has some embedded electronics (e.g., DS18S20 from Maxim).

Analogue sensors are usually measured by means of electronic interfaces similar to that shown in Fig. 2.1a. However, as explained throughout this chapter, analogue sensors can also be measured using the circuit topology proposed in Fig. 2.1b whenever the digital system excites them correctly. Actually, an appropriate excitation can convert analogue sensors to quasi-



2.2 Output signal of (a) an analogue sensor, (b) a quasi-digital sensor and (c) a digital sensor.

digital sensors since, as explained later in Section 2.4, the resulting output signal of the sensor is time modulated. Both quasi-digital and digital sensors can also be measured using the circuit topology shown in Fig. 2.1b, but the excitation signal coming from the digital system is generally not required; for the measurement of quasi-digital sensors, the digital system needs at least an embedded digital timer or counter.

Those analogue sensors that can be measured using direct interface circuits are described next; for more details, please refer to Baxter (1997), Pallàs-Areny and Webster (2001), and Fraden (2004). The explanation focuses on resistive and capacitive sensors, and discusses their topology, equivalent circuit model and applications.

2.2.1 Resistive sensors

Resistive sensors transform signals from a given energy domain to the electrical domain by changing its electrical resistance, which can be modelled as:

$$R = \rho \frac{l}{A} \qquad [2.1]$$

where ρ is the resistivity of the material, and l and A are the length and cross-section area of the conductor, respectively; note that any of the three parameters involved in Equation [2.1] can be altered by the measurand. The following subsections classify resistive sensors in terms of the number of sensing elements that make up the sensor and how these are interconnected.

Single resistive sensor

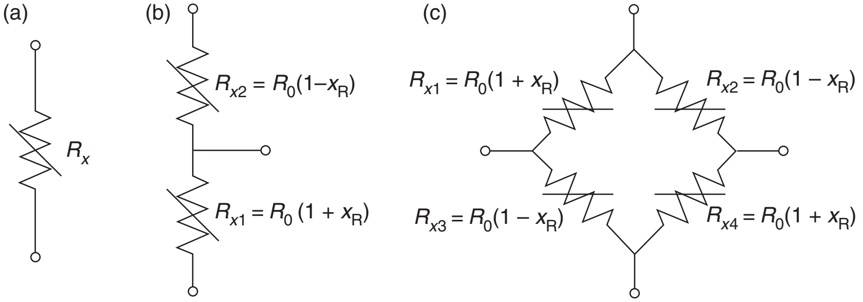
As shown in Fig. 2.3a, a single resistive sensor has one sensing element whose resistance (R_x) changes with the measurand. Such a resistance can be modelled as

$$R_x = R_0 + \Delta R = R_0(1 + x_R) \tag{2.2}$$

where R_0 is the nominal resistance at a reference value of the measurand, ΔR is the change of resistance due to (and, for some sensors, proportional to) the measurand and x_R is the relative change of resistance (i.e., $x_R = \Delta R/R_0$) due to the measurand. These sensors are commonly used to measure temperature (e.g., platinum sensors and thermistors), light (e.g., light-dependent resistors, LDR), gas (e.g., tin dioxide gas sensors) and humidity. Among the previous examples, only platinum sensors show an R_x that changes linearly with temperature. The other sensors generally require linearization techniques in the analogue and/or digital domain.

Differential resistive sensor

As shown in Fig. 2.3b, a differential resistive sensor has two sensing elements (R_{x1} and R_{x2}) that share a terminal and undergo opposite changes;



2.3 Resistive sensor with (a) single topology, (b) differential topology and (c) bridge topology.

that is, if R_{x1} increases with the measurand, then R_{x2} decreases and vice versa. Generally, x_R is assumed to be equal in magnitude but opposite in direction for R_{x1} and R_{x2} and, hence, these can be modelled as:

$$R_{x1} = R_0(1 + x_R) \quad \text{and} \quad R_{x2} = R_0(1 - x_R) \qquad [2.3]$$

Since R_{x1} and R_{x2} can be also altered by undesired inputs such as temperature (thus causing multiplicative errors), the measurand information is not carried by R_{x1} and R_{x2} but by x_R, which can be calculated by:

$$x_R = \frac{R_{x1} - R_{x2}}{R_{x1} + R_{x2}} \qquad [2.4]$$

Such sensors are frequently applied to measure, for example, linear or angular position/displacement, pressure (e.g., sensors based on Bourdon tubes), liquid level (e.g., float-based sensors) and magnetic field. The first three examples rely on a potentiometric differential sensor; for a linear potentiometric sensor, x_R is proportional to the measurand.

Bridge-type resistive sensor

A Wheatstone bridge with one, two or four sensing elements results in a quarter-bridge, half-bridge or full-bridge sensor, respectively. For the full-bridge topology (see Fig. 2.3c), which is the most usual since it provides the highest sensitivity, the four sensing elements undergo that same x_R but with opposite signs so that:

$$R_{x1} = R_{x4} = R_0(1 + x_R) \quad \text{and} \quad R_{x2} = R_{x3} = R_0(1 - x_R) \qquad [2.5]$$

These sensors are commonly used to measure weight (e.g., load cells based on metal strain gauges), pressure (e.g., sensors based on semiconductor strain gauges) and magnetic field (e.g., anisotropic magnetoresistive sensors (AMR) and giant magnetoresistive sensors (GMR)). For the previous examples, x_R is usually considered to change linearly with the measurand, at least within a given measurement range.

2.2.2 Capacitive sensors

Capacitive sensors transform signals from a given energy domain to the electrical domain by changing its electrical capacitance, which, for a parallel-plate arrangement, can be modelled as:

$$C = \varepsilon \frac{S}{d} \qquad [2.6]$$

where ε is the electric permittivity of the dielectric between plates, S is the area of the plates and d is the distance between plates; again, any of the three parameters involved in Equation [2.6] can be altered by the measurand. The following subsections classify capacitive sensors according to the number of sensing elements and their interconnection.

Single capacitive sensor

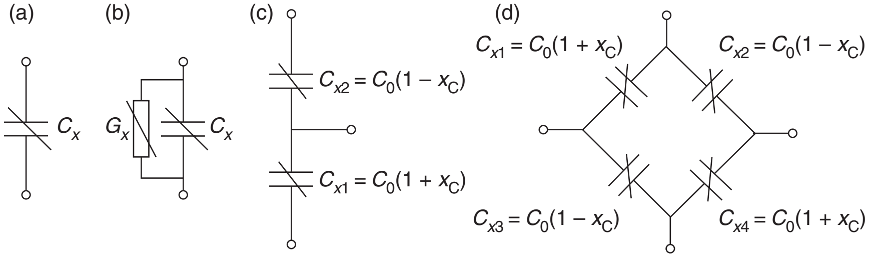
As shown in Fig. 2.4a, a single capacitive sensor has one sensing element whose capacitance (C_x) changes with the measurand. This capacitance can be modelled as:

$$C_x = C_0 + \Delta C = C_0(1 + x_C) \qquad [2.7]$$

where C_0 is the nominal capacitance at a reference value of the measurand, ΔC is the change of capacitance due to (and, for some sensors, proportional to) the measurand and x_C is the relative change of capacitance (i.e., $x_C = \Delta C/C_0$) due to the measurand. Such sensors are frequently applied to measure, for example, liquid level, humidity and gas.

Lossy capacitive sensor

Some capacitive sensors have a loss term that is usually modelled by a parasitic conductance G_x in parallel with C_x, as shown in Fig. 2.4b. Regrettably, the value of G_x is usually not constant, but it changes with the measurand, depends on environmental factors (such as temperature, condensation and pollution) and/or drifts with time. Capacitive sensors intended for the mea-



2.4 Capacitive sensor with (a) single topology, (b) single topology including a loss term, (c) differential topology and (d) bridge topology.

surement of proximity, humidity and two-component fluids concentration are common examples in which G_x can play an important role.

Differential capacitive sensor

As shown in Fig. 2.4c, a differential capacitive sensor has two sensing elements (C_{x1} and C_{x2}) that share an electrode and undergo opposite changes; that is, if C_{x1} increases with the measurand, then C_{x2} decreases and vice versa. Usually, x_C is considered equal but opposite for C_{x1} and C_{x2} and, hence, these can be modelled as:

$$C_{x1} = C_0(1 + x_C) \quad \text{and} \quad C_{x2} = C_0(1 - x_C) \qquad [2.8]$$

Note that Equation [2.8] is also valid when d is the parameter altered by the measurand in Equation [2.6] whenever the relative change of d is small. As in the resistive counterparts, x_C carries the measurand information and can be calculated by:

$$x_C = \frac{C_{x1} - C_{x2}}{C_{x1} + C_{x2}} \qquad [2.9]$$

These sensors are commonly used to measure linear or angular position/displacement, acceleration, tilt and pressure.

Bridge-type capacitive sensor

Sensing capacitive elements can also be interconnected in a bridge topology, as shown in Fig. 2.4d. In such a case, the four sensing elements undergo the same x_C but with opposite signs, so that:

$$C_{x1} = C_{x4} = C_0(1 + x_C) \quad \text{and} \quad C_{x2} = C_{x3} = C_0(1 - x_C) \qquad [2.10]$$

These sensors are frequently applied to measure, for instance, linear or angular position/displacement and pressure.

2.3 Microcontrollers

A microcontroller is a programmable processor-based digital IC widely used in control and measurement electronic systems. For the circuit shown in Fig. 2.1b, the μC will be the digital system in charge of exciting and measuring the sensor without using either the signal conditioning circuit or the ADC. Other digital systems, such as field programmable gate arrays (FPGA), have also been proposed to carry out such functions (Ares *et al.*, 2009; Vidal-Verdú *et al.*, 2011), but a µC seems more advisable because of its lower cost (say, $1) and lower power consumption (say, about 1 mA in active mode and less than 1 µA in power-down mode).

2.3.1 General description

Microcontrollers basically have three main blocks embedded into the same IC (Ivanov, 2008):

1. *Central processing unit (CPU):* The CPU is the core of the µC in charge of executing instructions sequentially (with an instruction cycle of T_S) to carry out a specific task. It incorporates an arithmetic logic unit (ALU) that processes the data (of 8, 16 or 32 bits) involved in such a task. The higher the number of bits, the higher the computational power, but also the higher the cost and power consumption. For this reason, direct interface circuits are generally implemented using an 8-bit µC. Some CPUs have a power-down (or sleep) mode that suspends their activity and, hence, decreases the power consumption, which is of interest for direct interface circuits.
2. *Memory*: The embedded memory saves the instructions to be executed and data to be processed. Most of the current commercial µCs have a Harvard architecture with two separate memories: an instructions (or program) memory and a data memory. The former is usually one-time programmable (OTP) or flash type, whereas the latter is Random Access Memory (RAM).
3. *Peripherals*: Peripherals are embedded hardware resources that enable the µC to interact with the off-chip world. They are controlled by the CPU but they operate in parallel with the CPU. Common peripherals integrated in current commercial µCs are: input/output ports, timers/counters, serial communication ports (e.g., UART, I2C, SPI), ADC, among others. Input/output ports and timers are essential peripherals to carry out the measurement in direct interface circuits.

Nowadays, there are many commercial µCs from different manufacturers but with quite similar features. Some examples are: the PIC16F family from Microchip Technology, the MSP430 family from Texas Instruments, the AVR family from Atmel, the HC08 family from Freescale, and the STM8 family from STMicroelectronics. As shown later in Section 2.5, direct interface circuits have been implemented using different commercial µCs, but the performance seems to be fairly independent of the µC used.

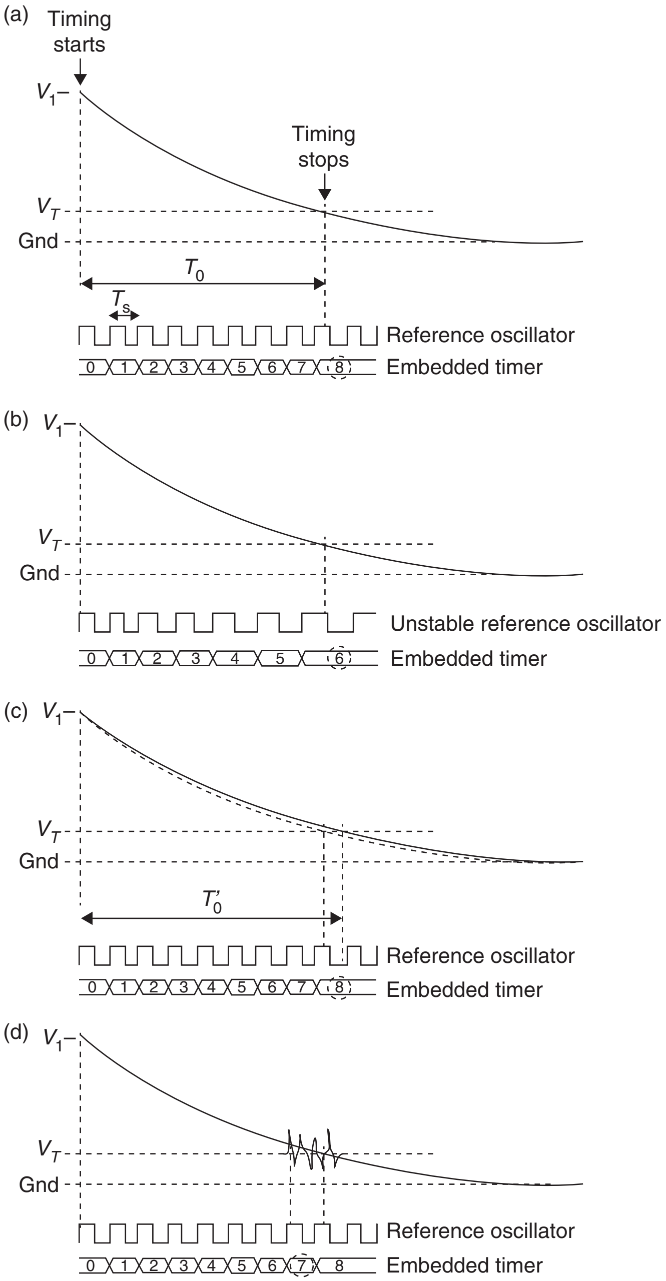
2.3.2 Time-interval measurement

An appropriate excitation signal in Fig. 2.1b can convert analogue sensors (such as resistive or capacitive sensors) to quasi-digital sensors providing a time-modulated signal. However, as explained later in Section 2.4, such a signal usually has a slow slew rate (SR); that is, the transition from '1' to '0' or vice versa is slow. For this reason, we next explain the operating principle of μCs when measuring such signals, the uncertainty sources and some design solutions to improve the measurement.

Let us suppose that the μC has to measure the time interval T_0 of the input signal shown in Fig. 2.5a, which has a slow SR; T_0 is actually the time-interval required to transition from V_1 (which is a high voltage identified as a digital '1') to V_T (which is an internal threshold voltage of a input port pin of the μC that determines the transition from '1' to '0'). In such a case, moreover, the start of the measurement is synchronized with the embedded timer, which is what really happens in the direct interface circuits described in Section 2.4. Once the measurement starts, the timer increases by 1 at every rising (or falling) edge of its reference oscillator whose period equals T_S; it is assumed that such a period equals the instruction cycle of the CPU, but they could be different using a prescaler. Then, when the input signal crosses V_T, the CPU issues a command to stop the time-interval measurement and to read the timer. In Fig. 2.5a, the measurement result is the digital number 8.

The time-to-digital conversion explained previously has three main uncertainty sources (Bouwens, 1984):

1. Instability of the reference oscillator: Since the measurement method uses an oscillator as a reference, an unstable oscillator brings about an unstable measurement result. For example, as shown in Fig. 2.5b, T_0 is converted to the digital number 6 instead of 8 when using an unstable reference oscillator. The main instability factors are time-drift, temperature dependence and sensitivity to supply voltage variations.
2. Quantization: The time-to-digital conversion suffers from quantization effects similar to those obtained in ADCs. For instance, if the time interval to be measured increases from T_0 to T_0', as shown in Fig. 2.5c, the resulting digital number is the same as before (i.e., 8) since the increase of the time interval is lower than T_S.
3. Trigger noise: The voltage comparison between the input signal and V_T is a noise-sensitive process that can be erroneously triggered. For example, in Fig. 2.5d, the voltage comparison is triggered in advance due to


(a)
Timing starts
V_1
Timing stops
V_T
Gnd
T_0
T_s
Reference oscillator
0 1 2 3 4 5 6 7 8
Embedded timer
(b)
V_1
V_T
Gnd
Unstable reference oscillator
0 1 2 3 4 5 6
Embedded timer
(c)
V_1
V_T
Gnd
T'_0
Reference oscillator
0 1 2 3 4 5 6 7 8
Embedded timer
(d)
V_1
V_T
Gnd
Reference oscillator
0 1 2 3 4 5 6 7 8
Embedded timer

2.5 (a) Time-interval measurement carried out by a μC, (b) effects of an unstable reference oscillator, (c) effects of quantization, (d) effects of trigger noise.

noise and the result is the digital number 7 instead of 8. Trigger noise has two components: external trigger noise (i.e., noise superimposed on the input signal to be measured) and internal trigger noise (i.e., noise superimposed on V_T). The latter arises from at least three different sources (Reverter and Pallàs-Areny, 2006): inherent (thermal) noise, power supply noise and CPU-activity noise.

Several design guidelines can be applied to reduce the effects of the previous uncertainty sources and, hence, to improve the time-interval measurement (Reverter and Pallàs-Areny, 2006). The main design rules for each uncertainty source are the following:

1. Microcontrollers can use three types of reference oscillator: an RC oscillator, ceramic resonator or crystal oscillator, the common specifications of which are summarized in Table 2.1. According to this table, accurate and stable time-interval measurements should be performed using a crystal oscillator as a reference, although it is the most expensive option. When using a crystal oscillator, the effects of its instability are generally much lower than those due to quantization and trigger noise.
2. Quantization error can be decreased by increasing the frequency of the reference oscillator of the timer, which nowadays can be up to tens of MHz. However, increasing such a frequency generally results in higher power consumption. Quantization error also depends on the technique and resources used by the CPU for detecting the voltage threshold crossing at the end of the measurement. Three alternatives can be distinguished:
   - Polling: The input signal to be measured is connected to a common input port pin whose state is periodically (every P instruction cycles) polled by the CPU. When the CPU detects a change (i.e., transition from '1' to '0' in Fig. 2.5a), it issues a command to read the timer. Accordingly, the effective quantization error ranges from 0 to $P \cdot T_S$ and, hence, it depends on the length of the polling loop.
   - General-purpose interrupt: The input signal is connected to a general-purpose external interrupt pin. When the voltage threshold is crossed, the CPU completes the execution of the current instruction and then executes the interrupt service routine (ISR) that reads the timer. However, because not all the instructions have the same execution time, the interrupt response time depends on the instruction being executed when the interrupt request arrives. Consequently, the effective

Table 2.1 Common specifications of oscillators used as a reference in microcontrollers

Oscillator	Inaccuracy (%)	Temperature coefficient (ppm/°C)	Time drifts (ppm/year)
RC oscillator	10	1000	–
Ceramic resonator	0.5	10	3000
Crystal oscillator	0.005	1	10

quantization error ranges from 0 to $M{\cdot}T_S$, where M is the number of instruction cycles needed to execute the longest instruction;

- Capture module: The input signal is connected to an input port pin associated to a capture module. This pin works as an external interrupt pin, but the current value of the timer is automatically captured by a special register when the voltage threshold is crossed. Then, once the current instruction has been executed, the CPU jumps to the ISR and reads the special register, not the timer. Therefore, the quantization error is independent of the instruction being executed and ranges from $-T_S$ to 0, which is the lowest one. Accordingly, a capture module of this type should be used for a high-resolution, accurate time-interval measurement.

3. Since noise superimposed on the supply voltage of the µC couples to V_T, the supply voltage should be as clean as possible. For this reason, it is highly recommended to have an appropriate decoupling capacitor between the power supply pins and to make a suitable layout of the ground and supply tracks accommodates sensitivity to noise. For input port pins with a Schmitt trigger (ST) buffer, noise coupling to the lower threshold voltage V_{TL} (which detects transitions from '1' to '0') is less than that coupling to the higher threshold voltage V_{TH} (which detects transitions from '0' to '1') and, hence, it is preferable to use V_{TL} for the voltage comparison (Reverter *et al.*, 2003). Furthermore, since the execution of instructions (especially those that require power, such as *jump* and *call* instructions) also brings about noise that couples to V_T, the activity of the CPU should be as low as possible while it waits for the voltage threshold to be crossed. Accordingly, it is suggested to place the CPU into sleep mode (whenever the timer and the interrupt system keep working) in order to decrease internal trigger noise. Note that if the CPU is placed into sleep mode and a general-purpose interrupt is used, then $M = 1$ and the resulting quantization error is equal (in absolute value) to that obtained using the capture module.

The previous guidelines for avoiding/reducing such uncertainty sources in time-interval measurements will be applied in the direct interface circuits proposed in the following sections.

2.4 Interface circuits

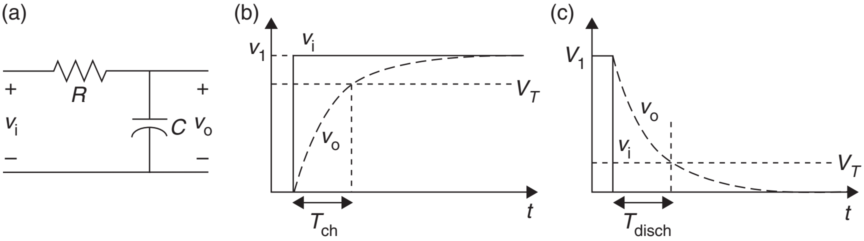
After describing sensors and µCs, this section explains how to join them to build a direct interface circuit. The operating principle of such circuits is described and, then, an explanation is offered on how to apply such an operating principle to measure each of the sensor topologies indicated in Sections 2.2.1 and 2.2.2.

2.4.1 Operating principle

Two measurement methods have been proposed for building the direct interface circuit topology shown in Fig. 2.1b:

1. Direct interfaces based on an RC circuit (Reverter and Pallàs-Areny, 2005a), where the µC measures the time interval needed to charge (or discharge) a capacitance *C* to a given threshold voltage through a resistance *R*. This method has been applied to measure resistive and capacitive sensors.
2. Direct interfaces based on charge transfer (Gaitán-Pitre *et al.*, 2009), where the µC counts the number of charge-transfer cycles needed to charge a reference capacitor to a given threshold voltage via a capacitive sensor. This method has been applied so far only to measure capacitive sensors.

This chapter focuses on direct interfaces based on an RC circuit (see Fig. 2.6a) operating as follows. If *C* is initially discharged and a step of amplitude V_1 is applied at the input, the output voltage is (see Fig. 2.6b)



2.6 (a) RC circuit, (b) charge process of an RC circuit, (c) discharge process of an RC circuit.

$$v_o(t) = V_1\left(1 - e^{-\frac{t}{RC}}\right) \tag{2.11}$$

and the time required to charge C from 0 to a given threshold voltage (V_T) is

$$T_{\text{ch}} = RC\ln\left(\frac{V_1}{V_1 - V_T}\right) \tag{2.12}$$

which is proportional to R and C. On the other hand, if C is already charged to V_1 and a step towards ground is applied at the input, the output voltage is (see Fig. 2.6c):

$$v_o(t) = V_1 e^{-\frac{t}{RC}} \tag{2.13}$$

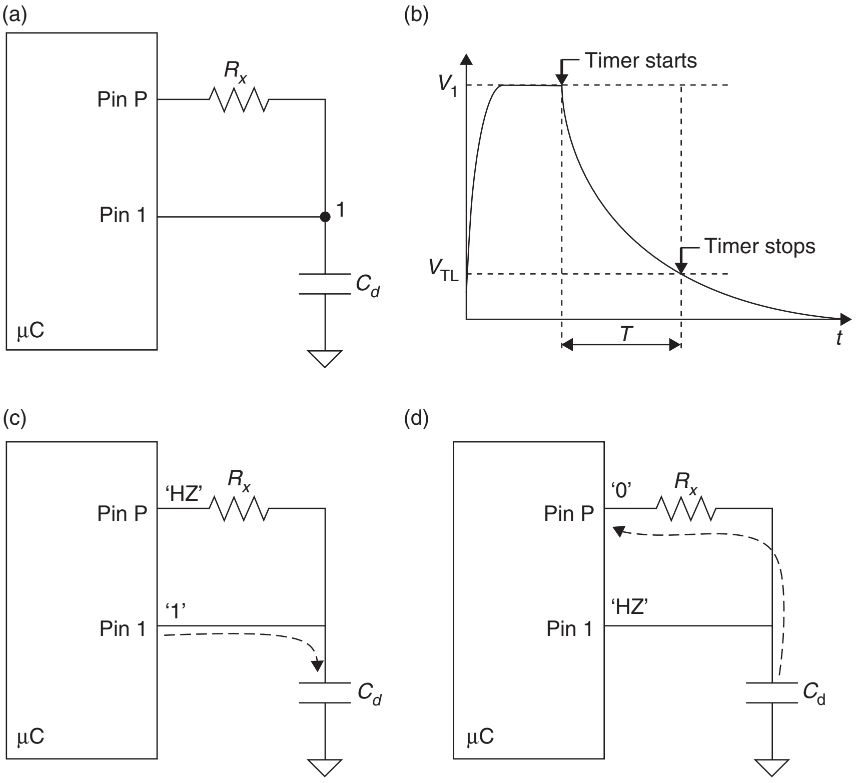
and the time needed to discharge C from V_1 to V_T is:

$$T_{\text{disch}} = RC\ln\left(\frac{V_1}{V_T}\right) \tag{2.14}$$

which, again, is proportional to R and C.

The RC circuit in Fig. 2.6a can be directly connected to a μC using the circuit topologies shown in Figs. 2.7a and 2.8a, where Pins 1 and P are two input/output digital port pins. Pin 1, which is in charge of monitoring the exponential charging or discharging voltage, usually includes a ST buffer (with two threshold voltages, V_{TL} and V_{TH}) and should be linked to a capture module (see Section 2.3.2). The circuits in Figs. 2.7a and 2.8a can measure either the charge time or the discharge time of the RC circuit, but it is expected that there will be less variability in the discharge-time measurement. This is because the discharge-time measurement uses V_{TL} as a threshold voltage, which is less noisy (see Section 2.3.2) than the V_{TH} used for the charge-time measurement. In both circuits (Figs. 2.7a and 2.8a), the measurement of the discharge time involves two operational steps: (1) charge, and (2) discharge and measurement.

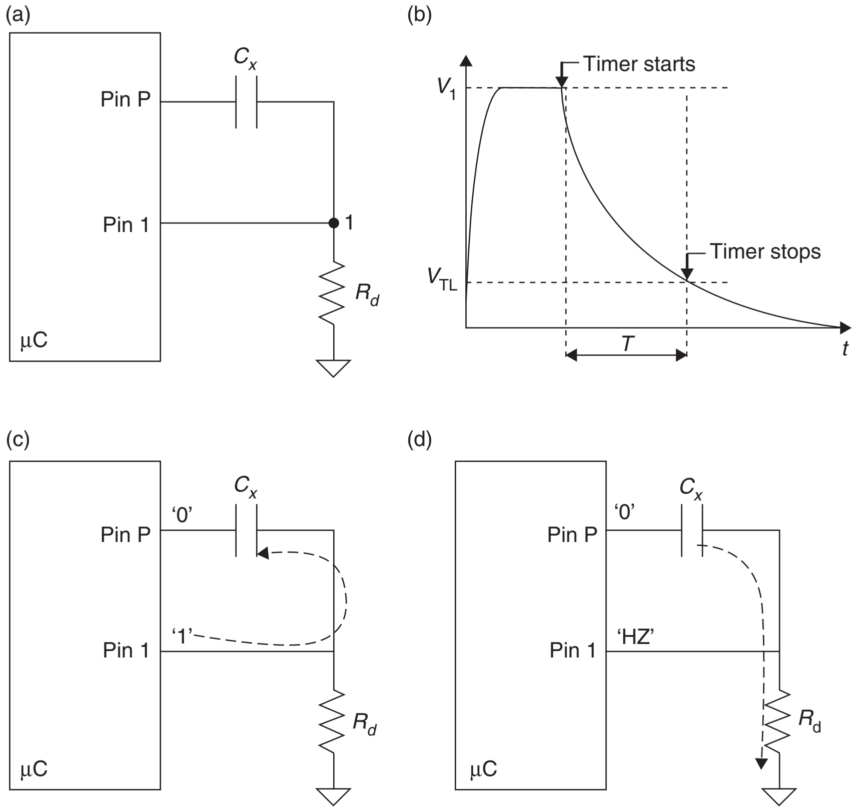
The circuit in Fig. 2.7a, which is intended for the measurement of a resistive sensor (R_x), operates as follows. During the charge stage (see Fig. 2.7c), Pin 1 is set as an output providing a digital '1', whereas Pin P is set as an input offering high impedance (HZ). Therefore, the capacitor C_d is rapidly



2.7 (a) Basic direct interface circuit for a resistive sensor (R_x), (b) waveform of the voltage across C_d during the charge–discharge process, (c) pin configuration during the charge stage, (d) pin configuration during the discharge stage.

charged to the analogue output voltage (V_1) corresponding to a digital '1', which is generally equal to the supply voltage (V_{DD}). During the discharge and measurement stage (see Fig. 2.7d), Pin 1 is set as an HZ input and Pin P is set as an output providing a digital '0'; consequently, C_d is discharged towards ground through R_x while the embedded timer measures the time interval required to do so. When the exponential discharging voltage crosses V_{TL}, the timer is read and a digital number proportional to R_x (see Equation [2.14]) is achieved. The resulting waveform of the voltage across C_d during the charge–discharge process is shown in Fig. 2.7b.

For the measurement of a capacitive sensor (C_x), it is recommended that the position of the resistance and the capacitance are exchanged, as shown in Fig. 2.8a. During the charge stage (see Fig. 2.8c), both pins are set as an output: Pin 1 provides a digital '1', whereas Pin P provides a



2.8 (a) Basic direct interface circuit for a capacitive sensor (C_x), (b) waveform of the voltage across R_d during the charge–discharge process, (c) pin configuration during the charge stage, (d) pin configuration during the discharge stage.

digital '0'. Therefore, C_x is rapidly charged to V_1. During the discharge and measurement stage (see Fig. 2.8d), Pin 1 is set as an HZ input and Pin P continues to provide a digital '0'; consequently, C_x is discharged towards ground through the resistor R_d while the timer is running. When the CPU detects that the voltage threshold has been crossed (see Fig. 2.8b), the timer is read and a digital number proportional to C_x (see Equation [2.14]) is achieved.

2.4.2 Circuits for resistive sensors

The operating principle explained in Fig. 2.7 can be applied to measure the three topologies of resistive sensor described in Section 2.2.1. Before discussing the proposed circuit for each sensor topology, a few general remarks are given on the components required:

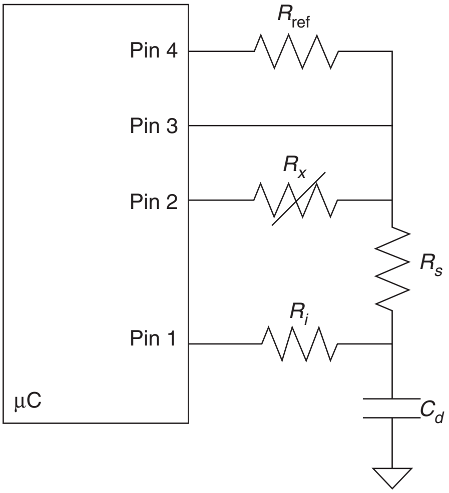
- The capacitor C_d is determined by a speed-resolution trade-off (Reverter and Pallàs-Areny, 2004), but for $T_S \approx 250$ ns it is advisable to operate with a time constant ($R{\cdot}C$) of about a few units of millisecond. For instance, the measurement of a resistive sensor of about 1000 Ω should use a C_d of a few units of microfarad (e.g., 2.2 μF).
- It is recommended that an additional resistor R_i is used between Pin 1 and Node 1, which improves the rejection of power supply noise/interference (Reverter *et al.*, 2007) but, however, at the expense of a longer charge stage. The cut-off frequency of the low-pass filter determined by R_i and C_d (during the charge stage) should be as low as possible but with a reasonable length of charging process (say, less than 1 ms). For example, if $C_d = 2.2$ μF then $R_i < 100$ Ω.
- It is also advisable to use an additional resistor R_s between Node 1 and R_x, which ensures that the discharging current is less than the maximum output current (I_max) sunk by a port pin even when R_x is very low. Assuming $V_\mathrm{DD} = 5$ V and $I_\mathrm{max} = 25$ mA, then $R_\mathrm{s} > 200$ Ω.

Single resistive sensor

The single resistive sensor shown in Fig. 2.3a can be directly connected to a μC using the circuit shown in Fig. 2.9 (Reverter *et al.*, 2005b). This circuit applies the three-signal auto-calibration technique (Meijer, 2008) to have a measurement result that is insensitive both to multiplicative and to additive parameters of the circuit. Accordingly, the interface circuit performs three discharge-time measurements:

1. *Sensor measurement*, which is intended to measure R_x;
2. *Reference measurement*, which is intended to measure a reference resistor (R_ref);
3. *Offset measurement*, which is intended to measure the offset brought about by the internal resistance (R_pin) of the port pins of the μC. Such resistance is assumed here to be the same for all the port pins but actually there is a mismatch of a few tenths of an ohm that generates offset and gain errors (Reverter *et al.*, 2005b).

Table 2.2 summarizes the state of Pins 2, 3 and 4 during the discharge stage and the resulting discharge time for each of the three measurements, where $k_R = C_d \ln\left(V_1/V_\mathrm{TL}\right)$. Once we have the three discharge times (T_x, T_ref and T_off), the sensor resistance can be estimated by:



2.9 Direct interface circuit for a single resistive sensor.

Table 2.2 Pins configuration and discharge time for each of the measurements carried out by the circuit in Fig. 2.9

Measurement	Pin 2	Pin 3	Pin 4	Discharge time
Sensor	'0'	HZ	HZ	$T_x = k_R(R_s + R_x + R_{pin})$
Reference	HZ	HZ	'0'	$T_{ref} = k_R(R_s + R_{ref} + R_{pin})$
Offset	HZ	'0'	HZ	$T_{off} = k_R(R_s + R_{pin})$

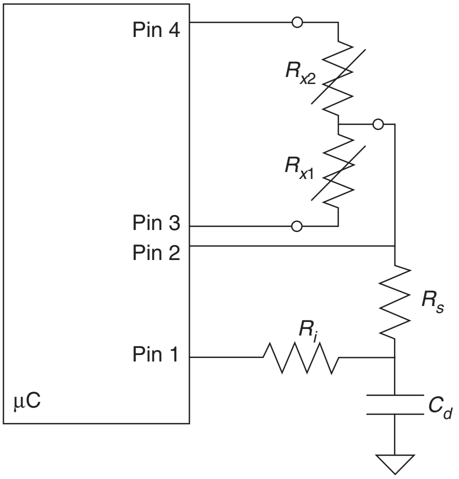
$$R_x = \frac{T_x - T_{off}}{T_{ref} - T_{off}} R_{ref} \qquad [2.15]$$

which is insensitive to the tolerance and low-frequency variability of C_d, V_1 and V_{TL}.

Differential resistive sensor

The direct interface circuit proposed for the measurement of differential resistive sensors (see Fig. 2.3b) is shown in Fig. 2.10 (Reverter and Casas, 2009a), which also performs three discharge-time measurements:

- *Sensor measurement #1*, which is intended to measure R_{x1}.
- *Sensor measurement #2*, which is intended to measure R_{x2}.
- *Offset measurement*, which is intended to measure R_{pin}.



2.10 Direct interface circuit for a differential resistive sensor.

Table 2.3 Pins configuration and discharge time for each of the measurements carried out by the circuit in Fig. 2.10

Measurement	Pin 2	Pin 3	Pin 4	Discharge time
Sensor #1	HZ	'0'	HZ	$T_1 = k_R(R_s + R_{x1} + R_{\text{pin}})$
Sensor #2	HZ	HZ	'0'	$T_2 = k_R(R_s + R_{x2} + R_{\text{pin}})$
Offset	'0'	HZ	HZ	$T_{\text{off}} = k_R(R_s + R_{\text{pin}})$

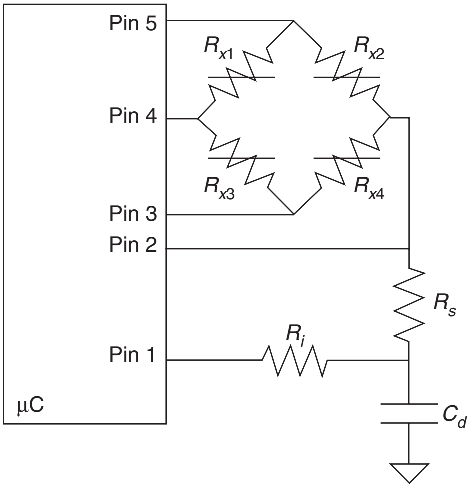
The state of Pins 2, 3 and 4 during the discharge stage and the discharge time for each of the three measurements is summarized in Table 2.3. By means of these three discharge times (T_1, T_2 and T_{off}), the parameter x_R of the differential sensor can be estimated by:

$$x_R = \frac{T_1 - T_2}{T_1 + T_2 - 2T_{\text{off}}} \qquad [2.16]$$

which is similar to Equation [2.4] but, instead of resistances, it uses discharge times. Unlike the measurement of single resistive sensors, x_R can be estimated here without using any reference resistor.

Bridge-type resistive sensor

Resistive sensors in a bridge topology (see Fig. 2.3c) can be directly connected to a μC using the interface circuit shown in Fig. 2.11 (Sifuentes *et al.*, 2008). This circuit measures four discharge times (T_1, T_2, T_3 and T_{off}) by



2.11 Direct interface circuit for a bridge-type resistive sensor.

applying the configuration of pins indicated in Table 2.4. Accordingly, for a full-bridge topology, the parameter x_R of the sensor can be estimated by:

$$x_R = \frac{T_1 - T_3}{T_2 - T_{\text{off}}} \qquad [2.17]$$

For other bridge topologies, x_R can be estimated using other time-based equations (Sifuentes *et al.*, 2008). Furthermore, for sensors whose output is temperature dependent (e.g., piezoresistive pressure sensors), the result obtained from Equation [2.17] can be easily corrected by estimating the temperature by means of the sensor itself (Reverter *et al.*, 2009b).

2.4.3 Circuits for capacitive sensors

The four topologies of capacitive sensor described in Section 2.2.2 can be directly measured by a μC using the operating principle explained in Fig. 2.8. Again, before discussing the proposed circuits, a few general remarks are given on the components required:

- The resistor R_d is determined by a speed-resolution trade-off (Reverter and Pallàs-Areny, 2004). However, unlike Section 2.4.2, here it is advisable to operate with a shorter time constant (say, a few hundreds of microsecond). Otherwise the required R_d is too high and Node 1 (see Fig. 2.8a) becomes a high-impedance point, this increasing the susceptibility to

Table 2.4 Pins configuration and discharge time for each of the measurements carried out by the circuit in Fig. 2.11

Measurement	Pin 2	Pin 3	Pin 4	Pin 5	Discharge time
Sensor #1	HZ	'0'	HZ	HZ	$T_1 = k_R[R_s + (R_{x4}\|\|(R_{x1} + R_{x2} + R_{x3}))+ R_{\text{pin}}]$
Sensor #2	HZ	HZ	'0'	HZ	$T_2 = k_R[R_s+ ((R_{x3} + R_{x4})\|\|(R_{x1} + R_{x2}))+ R_{\text{pin}}]$
Sensor #3	HZ	HZ	HZ	'0'	$T_3 = k_R[R_s + (R_{x2}\|\|(R_{x1} + R_{x3} + R_{x4}))+ R_{\text{pin}}]$
Offset	'0'	HZ	HZ	HZ	$T_{\text{off}} = k_R(R_s + R_{\text{pin}})$

Note: $R_A \,||\, R_B = \dfrac{(R_A \cdot R_B)}{(R_A + R_B)}$.

interference, Moreover, the effects of G_x (see Fig. 2.4b) would be more significant when using a high value of R_d. For instance, the measurement of a capacitive sensor of 150 pF should use an R_d of about a few units of megaohm (e.g., 1 MΩ);

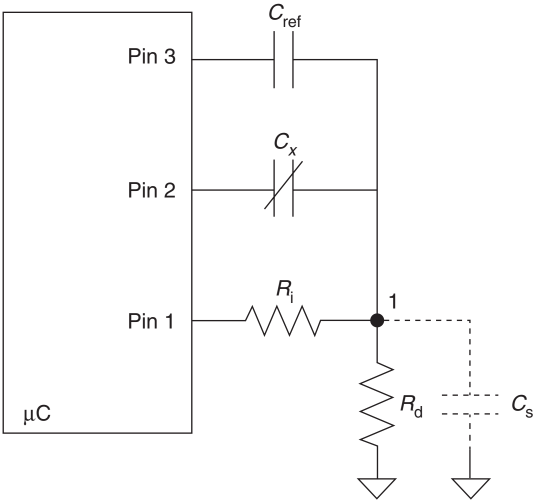
- It is also recommended that an additional resistor (R_i) be used between Pin 1 and Node 1 to improve the rejection of power supply noise/interference (Reverter *et al.*, 2007). The cut-off frequency of the low-pass filter determined by R_i and C_x (during the charge stage) should be as low as possible but, at the same time, R_i must be considerably less than R_d in order to ensure an appropriate charge of C_x. For example, if $R_d = 1$ MΩ then $R_i < 1$ kΩ.

Single capacitive sensor

The single capacitive sensor shown in Fig. 2.4a can be directly connected to a µC using the circuit shown in Fig. 2.12 (Reverter and Casas, 2008a). As in its resistive counterpart, the circuit applies the three-signal auto-calibration technique (Meijer, 2008) and, for this reason, it carries out three discharge-time measurements:

- *Sensor measurement*, which is intended to measure C_x.
- *Reference measurement*, which is intended to measure a reference capacitor (C_{ref}).
- *Offset measurement*, which is intended to measure the offset due to the parasitic capacitance (C_s) between Node 1 and ground. The parasitic capacitances of the port pins set as an HZ input are assumed negligible throughout Section 2.4.3, but their effects are carefully analysed in (Reverter and Casas, 2010a).

Table 2.5 summarizes the state of Pins 2 and 3 during the charge–discharge process and the discharge time for each of the three measurements, where



2.12 Direct interface circuit for a single capacitive sensor.

Table 2.5 Pins configuration and discharge time for each of the measurements carried out by the circuit in Fig. 2.12

Measurement	Pin 2	Pin 3	Discharge time
Sensor	'0'	HZ	$T_x = k_C(C_x + C_s)$
Reference	HZ	'0'	$T_{ref} = k_C(C_{ref} + C_s)$
Offset	HZ	HZ	$T_{off} = k_C C_s$

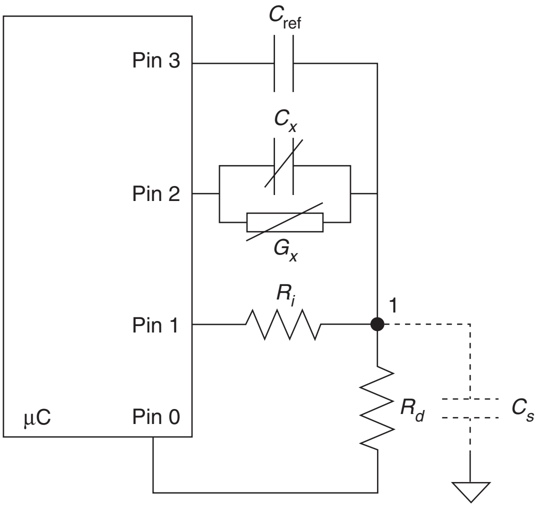
$k_C = R_d \ln(V_1/V_{TL})$. Once the three discharge times (T_x, T_{ref} and T_{off}) are measured, the sensor capacitance can be estimated by:

$$C_x = \frac{T_x - T_{off}}{T_{ref} - T_{off}} C_{ref} \qquad [2.18]$$

which, again, is insensitive to the tolerance and low-frequency variability of R_d, V_1 and V_{TL}.

Lossy capacitive sensor

If the circuit in Fig. 2.12 uses an extra port pin (Pin 0, in Fig. 2.13) and performs an additional discharge-time measurement T_{ad} (see Table 2.6, where



2.13 Direct interface circuit for a lossy capacitive sensor.

Table 2.6 Pins configuration and discharge time for each of the measurements carried out by the circuit in Fig. 2.13

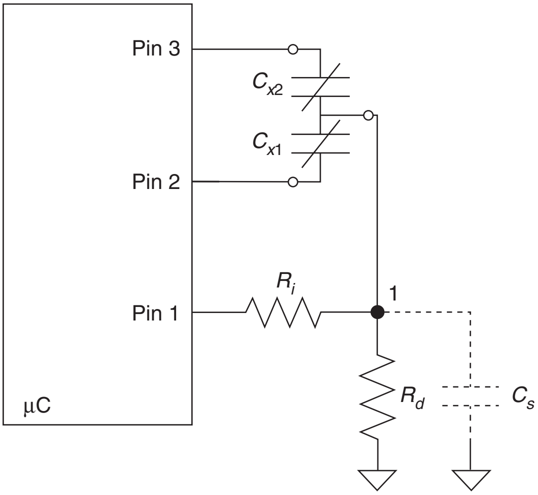
Measurement	Pin 0	Pin 2	Pin 3	Discharge time
Sensor	'0'	'0'	HZ	$T_x = k(R_d \| G_x^{-1})(C_x + C_s)$
Reference	'0'	HZ	'0'	$T_{ref} = kR_d(C_{ref} + C_s)$
Offset	'0'	HZ	HZ	$T_{off} = kR_dC_s$
Additional	HZ	'0'	HZ	$T_{ad} = kG_x^{-1}(C_x + C_s)$

$k = \ln(V_1/V_{TL})$), the capacitance and parasitic conductance of a lossy capacitive sensor (see Fig. 2.4b) can be estimated by (Reverter and Casas, 2010b):

$$C_x = \frac{T_x T_{off} + T_{ad}(T_x - T_{off})}{(T_{ad} - T_x)(T_{ref} - T_{off})} C_{ref} \qquad [2.19]$$

$$G_x = \frac{T_x}{T_{ad} - T_x} \frac{1}{R_d} \qquad [2.20]$$

Note that if there is a very low value of G_x, then T_{ad} is much longer than the other discharge times and, hence, Equation [2.19] can be simplified to Equation [2.18].



2.14 Direct interface circuit for a differential capacitive sensor.

Table 2.7 Pins configuration and discharge time for each of the measurements carried out by the circuit in Fig. 2.14

Measurement	Pin 2	Pin 3	Discharge time
Sensor #1	'0'	HZ	$T_1 = k_C(C_{x1} + C_s)$
Sensor #2	HZ	'0'	$T_2 = k_C(C_{x2} + C_s)$
Sensor #3	'0'	'0'	$T_3 = k_C(C_{x1} + C_{x2} + C_s)$

Differential capacitive sensor

The direct interface circuit proposed to measure differential capacitive sensors (see Fig. 2.4c) is shown in Fig. 2.14 (Reverter and Casas, 2010a), and performs the three discharge-time measurements indicated in Table 2.7. By means of these three discharge times (T_1, T_2 and T_3), x_C can be estimated by:

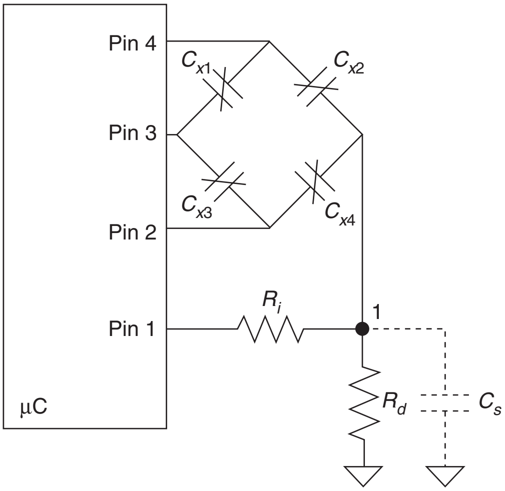
$$x_C = \frac{T_1 - T_2}{T_3} \qquad [2.21]$$

The parameter x_C can also be estimated by performing an offset measurement and applying a time-based equation similar to Equation [2.16] (Reverter and Casas, 2008b), but the result becomes more sensitive to the parasitic capacitances of the port pins set as an HZ input. On the other hand, as in its resistive counterpart, the proposed circuit does not require any reference capacitor.

Bridge-type capacitive sensor

Capacitive sensors in a bridge topology (see Fig. 2.4d) can also be directly connected to a µC using the interface circuit shown in Fig. 2.15. This circuit measures four discharge times (T_1, T_2, T_3 and T_{off}) by applying the configuration of pins indicated in Table 2.8. For a full-bridge topology, x_C can be estimated by:

$$x_C = \frac{T_1 - T_3}{T_2 - T_{\text{off}}} \qquad [2.22]$$



2.15 Direct interface circuit for a bridge-type capacitive sensor.

Table 2.8 Pins configuration and discharge time for each of the measurements carried out by the circuit in Fig. 2.15

Measurement	Pin 2	Pin 3	Pin 4	Discharge time
Sensor #1	'0'	'0'	HZ	$T_1 = k_C[(C_{x1} \oplus C_{x2}) + C_{x4} + C_s]$
Sensor #2	'0'	HZ	'0'	$T_2 = k_C[C_{x2} + C_{x4} + C_s]$
Sensor #3	HZ	'0'	'0'	$T_3 = k_C[C_{x2} + (C_{x3} \oplus C_{x4}) + C_s]$
Offset	HZ	HZ	HZ	$T_{\text{off}} = k_C C_s$

Note: $C_A \oplus C_B = \frac{(C_A \cdot C_B)}{(C_A + C_B)}$.

For other bridge topologies, x_C can be estimated using similar time-based equations.

2.5 Applications

This section describes how the direct interface circuits explained in Section 2.4 have been implemented using commercial low-cost μCs (see Table 2.9) and then applied to measure commercial sensors (see Table 2.10).

2.5.1 Temperature measurement

The direct interface circuit shown in Fig. 2.9 using μC #1 (see Table 2.9) was applied to measure resistors that emulated the Pt1000 temperature sensor

Table 2.9 Main features of the commercial microcontrollers used to implement the direct interface circuits

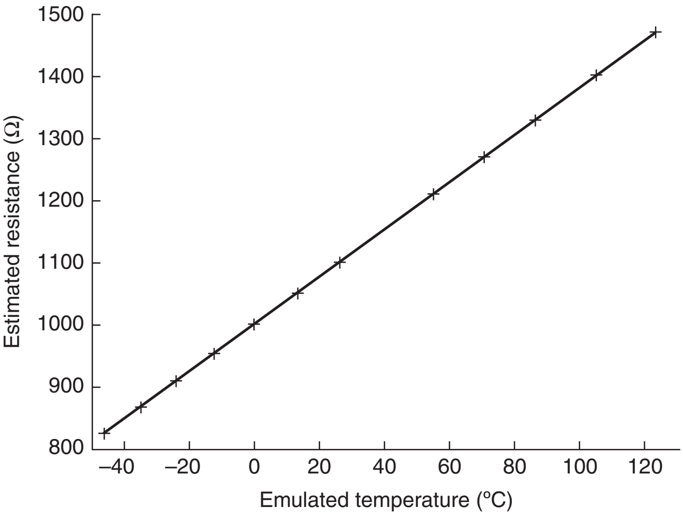
	μC #1	μC #2	μC #3
Manufacturer	Microchip	Atmel	Texas Instruments
Model	PIC16F873	AVR ATtiny2313	MSP430F123
Supply voltage (V_{DD})	5.0 V	5.0 V	3.0 V
Threshold voltage (V_{TL})	1.4 V	2.2 V	1.1 V
Reference oscillator	Crystal – 20 MHz	Crystal – 20 MHz	Crystal – 4 MHz
Embedded timer	16-bit Timer1	16-bit Timer1	16-bit Timer_A3
Time-base of timer (T_S)	200 ns	50 ns	250 ns
Crossing detection	External interrupt	Capture module	External interrupt
Sleep mode	No[a]	Yes	Yes

[a] The sleep mode was not used during the discharge-time measurement because it stops the time-base of the timer.

Table 2.10 Main features of the commercial sensors measured by the direct interface circuits

	Temperature sensor	Magnetic sensor	Humidity sensor	Accelerometer
Manufacturer	–	Honeywell	Humirel	VTI Technologies
Model	Pt1000	HMC1052	HS1101	SCG10Z-G001CC
Type	Resistive	Resistive (AMR)	Capacitive	Capacitive
Topology	Single	Full bridge	Single	Differential
Typical R_0 or C_0	1000 Ω at 0°C	1000 Ω at 0 μT	180 pF at 55%RH	1.5 pF at rest
Typical sensitivity	+3.81 Ω/°C	± 1 Ω/100 μT	+0.34 pF/%RH	±0.105 pF/g
Non-linearity	–	<1.8% FSS[a]	<2% FSS[a]	–
Tested range	[–45, 120]°C	[75, 600] μT	[10, 90]%RH	[–1, +1] g

[a] Measured values.

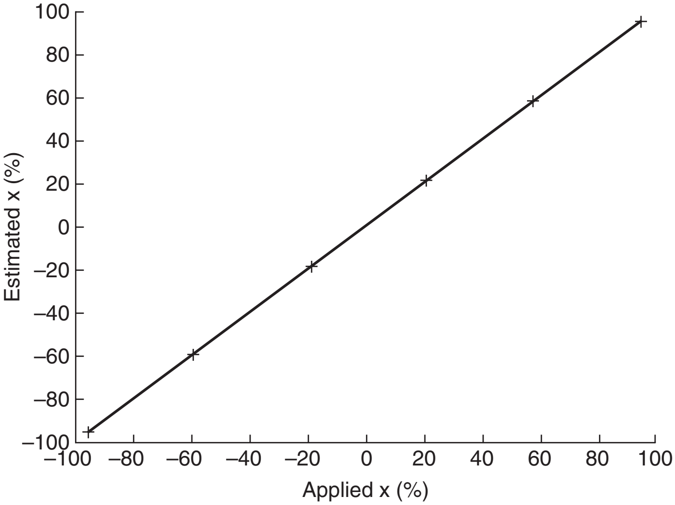


2.16 Experimental results of the circuit in Fig. 2.9 when measuring a temperature sensor.

indicated in Table 2.10 (Reverter *et al.*, 2005b); the other components of the circuit were $R_{\text{ref}} = 1470\ \Omega$, $R_s = 330\ \Omega$ and $C_d = 2.2\ \mu\text{F}$. Figure 2.16 shows the resistance estimated by Equation [2.15] for temperatures between –45°C and +120°C, which were emulated by resistors. According to the straight line fitted to the experimental data by means of the least-squares method, the maximum non-linearity error was 0.01% FSS. On the other hand, the effective resolution was almost 13 bits (which corresponds to 0.1 Ω and, hence, to 0.025°C) for a measuring time of about 50 ms required to average ten measurements. From the author's point of view, such values of linearity and resolution are quite remarkable, taking into account the simplicity of the proposed interface circuit.

2.5.2 Position measurement

A 1-kΩ linear potentiometer emulating a potentiometric position sensor was measured by the direct interface circuit in Fig. 2.10 using μC #2 (see Table 2.9) (Reverter and Casas, 2009a); the other components of the circuit were $R_s = 470\ \Omega$, $R_i = 100\ \Omega$ and $C_d = 470$ nF. Figure 2.17 shows the value of x_R estimated by Equation [2.16] for different positions of the wiper of the potentiometer. According to the fitted straight line, the maximum non-linearity error was 0.01% FSS, which was mainly due to quantization effects in the discharge-time measurement. On the other hand, the effective resolution was almost 13 bits (which corresponds to changes of x_R equal to

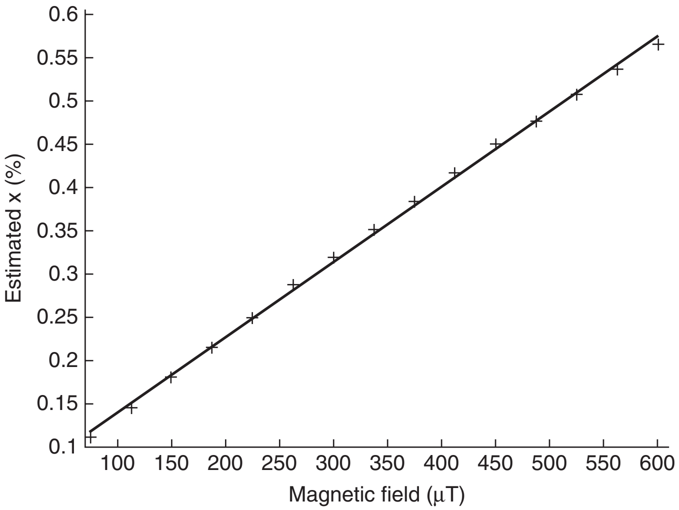


2.17 Experimental results of the circuit in Fig. 2.10 when measuring a potentiometric sensor.

0.025%) for a measuring time of about 100 ms required to average one hundred measurements. The performance in terms of linearity and resolution was quite similar to that obtained in Section 2.5.1.

2.5.3 Magnetic field measurement

The direct interface circuit shown in Fig. 2.11 using μC #3 (see Table 2.9) was applied to measure the magnetic sensor indicated in Table 2.10 (Sifuentes *et al.*, 2008); the other components of the circuit were R_i = 120 Ω and C_d = 2.2 μF. Figure 2.18 shows the value of x_R estimated by [2.17] for magnetic fields between 75 μT and 600 μT generated by a current-controlled solenoid; in fact, the sensor could operate between –600 μT and +600 μT, but it was tested in a lower range due to limitations of the measurement set-up. The estimated x_R was always positive and very small (lower than 0.6%), but it agrees with the typical sensitivity of the sensor (see Table 2.10). According to the fitted straight line, the maximum non-linearity error was 1.8% FSS, but this was mainly due to the non-linearity of the sensor (see Table 2.10); actually, when the circuit measured a bridge circuit emulated by resistors instead of the sensor, the maximum non-linearity error was about 0.1% FSS. Therefore, the sensor – not the interface circuit, in spite of its simplicity – limited the accuracy of the measurement, as is to be expected from a well-designed interface circuit. On the other hand, assuming the overall measurement range, the effective resolution

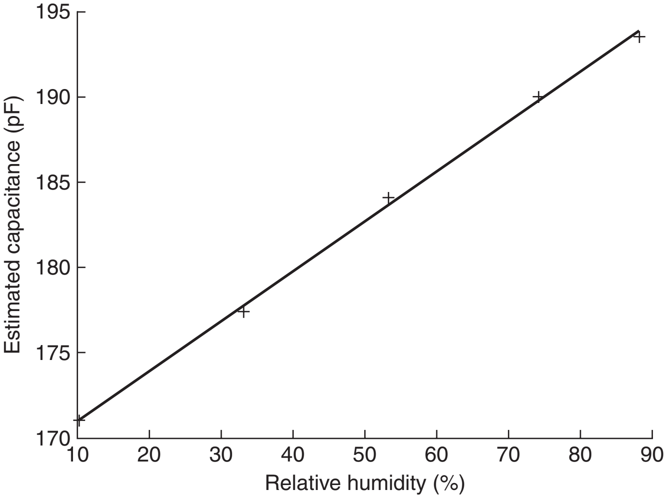


2.18 Experimental results of the circuit in Fig. 2.11 when measuring a magnetic sensor.

was about 7 bits (which corresponds to 10 μT) for a measuring time of about 50 ms required to average ten measurements. Note that for the same resolution in ohms (i.e., 0.1 Ω), the resolution in bits here is smaller than that in Section 2.5.1 since the overall change of resistance is smaller (i.e., ± 6 Ω according to Table 2.10).

2.5.4 Relative humidity measurement

The humidity sensor shown in Table 2.10 was measured by the direct interface circuit in Fig. 2.12 using μC #2 (see Table 2.9) (Reverter and Casas, 2008a); the other components of the circuit were $C_{\text{ref}} = 177$ pF, $R_d = 1$ MΩ and $R_i = 1$ kΩ. Figure 2.19 shows the capacitance estimated by Equation [2.18] for different values of relative humidity (RH) obtained by means of saturated salt solutions. The capacitance was between 171 pF and 194 pF for an RH range between 10% and 90%. The maximum non-linearity error was 2.0% FSS, but this was mainly due to the non-linearity of the sensor (see Table 2.10); in fact, when the circuit measured capacitors instead of the capacitive sensor, the maximum non-linearity error was 0.1% FSS. Therefore, as in Section 2.5.3, the sensor (and not the interface circuit) limited the accuracy of the measurement. The effective resolution was 9 bits (which corresponds to 50 fF and, hence, to 0.2% RH) for a measuring time of about 50 ms required to average one hundred measurements.



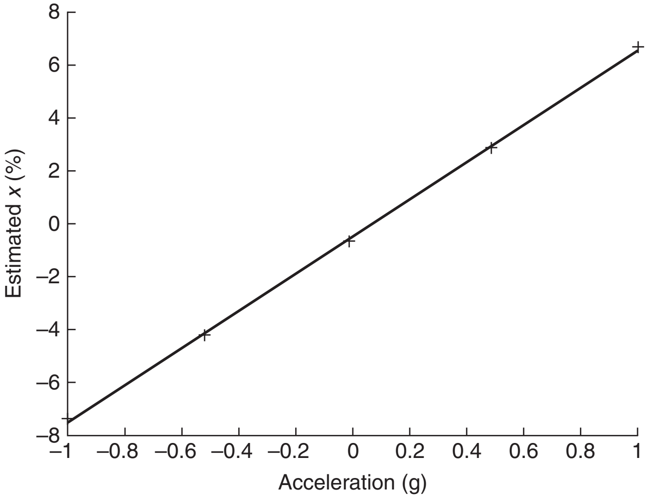
2.19 Experimental results of the circuit in Fig. 2.12 when measuring a humidity sensor.

2.5.5 Tilt measurement

The direct interface circuit in Fig. 2.14 using μC #2 (see Table 2.9) was applied to measure the accelerometer shown in Table 2.10 (Reverter and Casas, 2010a); the other components of the circuit were R_d = 20 MΩ and R_i = 10 kΩ. Figure 2.20 shows the value of x_C estimated by Equation [2.21] for accelerations between –1 g and +1 g, which correspond to inclinations between –90° and +90°. The estimated x_C was between –7% and +7%, which agrees with the typical sensitivity of the sensor (see Table 2.10). The maximum non-linearity error was 1.1% FSS, which was basically due to quantization effects in the discharge-time measurement. The effective resolution was 7 bits (which corresponds to 15 mg) for a measuring time of about 50 ms required to average ten measurements. In comparison with Section 2.5.4, the performance of the circuit in Fig. 2.14 was not as good, but this is because it had to measure a very low value capacitive sensor with a very low sensitivity (see Table 2.10).

2.5.6 Other applications

Direct interface circuits have also been applied to measure other physical and chemical quantities; for example: atmospheric pressure (Jordana and Pallàs-Areny, 2006), tactile pressure (Vidal-Verdú *et al.*, 2011), gas (Courbat *et al.*, 2012) and light (Sifuentes, 2009b).



2.20 Experimental results of the circuit in Fig. 2.14 when measuring an accelerometer.

2.6 Future trends

The field of direct interface circuits is still under research and many interesting ideas could be developed in the near future. From the author's point of view, future work on direct interface circuits could be focussed on:

- Applications: Many measurement systems based on resistive and capacitive sensors, but especially those intended for low-cost low-power applications, could benefit from the advantages of direct interface circuits. In fact, recently, such circuits have been proposed to measure: low-power magnetic sensors for vehicle detection (Sifuentes *et al.*, 2011), low-cost environmental sensors made by inkjet printing of silver on paper (Courbat *et al.*, 2011), low-power gas sensors (Courbat *et al.*, 2012) and low-cost low-power RH sensors to be integrated into radio frequency identification (RFID) labels (Pelegrí-Sebastiá *et al.*, 2012).
- Improving the current operating principle: The performance of direct interfaces based on an RC circuit could be improved by incorporating new ideas and/or techniques; for instance, the so-called Vernier technique (Sifuentes *et al.*, 2009a). Such techniques could either improve the resolution or reduce the cost, measuring time and power consumption while preserving the same resolution.
- Developing novel operating principles: As indicated in Section 2.4.1, direct interface circuits can also rely on the charge-transfer technique. Circuits based on such a technique have not been widely studied so

far (Dietz *et al.*, 2002; Gaitán-Pitre *et al.*, 2009), but their performance could be quite promising, especially when measuring low value capacitive sensors, thanks to better interference rejection. The measurement of resistive sensors using the charge-transfer technique could also be investigated.

- Measuring other sensors: Most of the research work done so far on direct interface circuits has been focussed on measuring resistive and capacitive sensors, but the direct measurement of other analogue sensors is also of interest. The design and analysis of direct interface circuits, for instance, for inductive sensors, current-output sensors (e.g., a photodiode) (Stojanovic and Karadaglic, 2007) and/or impedance sensors (Czaja, 2012) using common or novel operating principles could be an interesting topic of research.
- Using novel peripherals embedded into µCs: This field of technology is developing at a rapid rate. It almost seems that, with every passing day, commercial µCs are designed with more embedded resources and peripherals. Some of them – for example, analogue comparators or configurable logic cells – could be of interest to improve the circuit performance and/or to develop novel operating principles. The power-down modes included in novel commercial µCs could also be of help to decrease the power consumption of such circuits even further.
- Designing an ASIC or ASSP: The design of an application-specific integrated circuit (ASIC) or an application-specific standard processor (ASSP) could also be of interest in order to have a digital system optimized in terms of power and embedded resources. Actually, a commercial ASSP known as USTI (which means universal sensors and transducers interface) has been recently designed. This chip is able to measure resistive sensors (Yurish, 2009a, 2011) and capacitive sensors (Yurish, 2009b) using the operating principle explained in Section 2.4.

2.7 Sources of further information and advice

Researchers worldwide are involved in the field of direct interface circuits. Some of these researchers, their affiliation and a brief description of their work are indicated below.

- Courbat, J. and Briand, D. (Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland), who use direct interface circuits to measure low-cost low-power sensors.
- Czaja, Z. (Gdansk University of Technology, Gdansk, Poland), who proposes direct interface circuits for impedance sensors.

- Gaitán-Pitre, J.E. and Pallàs-Areny, R. (Universitat Politècnica de Catalunya, Castelldefels, Spain), who undertake research on direct interface circuits based on the charge-transfer technique.
- Kokolanski, Z. and Gavrovski, C. (University Ss. Cyril and Methodius, Skopje, Macedonia), who propose novel techniques to reduce the uncertainty of measurement (Kokolanski *et al.*, 2011).
- Pelegrí-Sebastiá, J. (Universidad Politécnica de Valencia, Gandia, Spain), who uses direct interface circuits to measure low-cost, low-power sensors.
- Sifuentes, E. (Universidad Autónoma de Ciudad Juárez, Juárez, Mexico), who undertakes research on direct interface circuits for vehicle detection.
- Vidal-Verdú, F. (University of Malaga, Malaga, Spain), who uses FPGAs to build direct interface circuits for tactile sensors.
- Yurish, S. (Technology Assistance BCN2010, Barcelona, Spain), who designs ASSPs based on the operating principle explained throughout this chapter.

2.8 References

Ares, L., Rodríguez-Andina, J.J. and Fariña, J. (2009) 'FPGA-based direct resistance and capacitance measurement', in *Proceedings IEEE IECON 2009*, Porto, Portugal, 3–5 November 2009, 2837–41.

Baxter, L.K. (1997) *Capacitive Sensors. Design and Applications*, New York, IEEE Press.

Bierl, L. (1996) *Precise measurements with the MSP430*, Dallas, TX: Texas Instruments.

Bouwens, A.J. (1984) *Digital instrumentation*, New York, McGraw-Hill.

Courbat, J., Kim, Y.B., Briand, D. and Rooij, N.F. (2011) 'Inkjet printing on paper for the realization of humidity and temperature sensors', in *Proceedings Transducers 2011*, Beijing, China, 5–7 June 2011, 1356–9.

Courbat, J., Briand, D., Yue, L., Raible, S. and Rooij, N.F. (2012) 'Drop-coated metal-oxide gas sensor on polyimide foil with reduced power consumption for wireless applications', *Sensors and Actuators B*, **161**, 862–8.

Cox, D. (1997) *Implementing Ohmmeter/Temperature Sensor*, AN512, Chandler, AZ: Microchip Technology Inc.

Czaja, Z. (2012) 'A microcontroller system for measurement of three independent components in impedance sensors using a single square pulse', *Sensors and Actuators A*, **173**, 284–92.

Dietz, P.H., Leigh, D. and Yerazunis, W.S. (2002), 'Wireless liquid level sensing for restaurant applications', in *Proceedings IEEE Sensors 2002*, Orlando, USA, 12–14 June 2002, 715–19.

Fraden, J. (2004) *Handbook of Modern Sensors. Physics, Designs, and Applications*, New York: Springer, 3rd edn

Gaitán-Pitre, J.E., Gasulla, M. and Pallàs-Areny, R. (2009) 'Analysis of a direct interface circuit for capacitive sensors', *IEEE Transactions on Instrumentation and Measurement*, **58** (9), 2931–7.

Huising, J.H. (2008) 'Smart sensor systems: Why? Where? How?', in Meijer, G.C.M (ed.) *Smart Sensor Systems*, Chichester (UK), Wiley, 1–21.

Ivanov, R.M. (2008) 'Microcontrollers and digital signal processors for smart sensors systems', in Meijer, G.C.M (ed.), *Smart Sensor Systems*, Chichester (UK), Wiley, 343–74.

Jordana, J. and Pallàs-Areny, R. (2006) 'A simple, efficient interface circuit for piezoresistive pressure sensors', *Sensors and Actuators A*, **127**, 69–73.

Kirianaki, N.V., Yurish, S.Y., Shpak, N.O. and Deynega, V.P. (2002) *Data Acquisition and Signal Processing for Smart Sensors*, Chichester (UK), John Wiley & Sons.

Kokolanski, Z., Gavrovski, C. and Dimcev, V. (2011) 'Continual one point auto-calibration technique in simple sensor-microcontroller interface', *Electronics*, **15** (1), 72–7.

Meijer, G.C.M. (2008) 'Interface electronics and measurement techniques for smart sensor sytems' in Meijer, G.C.M (ed.) *Smart Sensor Systems*, Chichester (UK), Wiley, 23–54.

Pallàs-Areny, R. and Webster, J.G. (2001) *Sensors and Signal Conditioning*, New York, John Wiley & Sons, 2nd edn.

Pelegrí-Sebastiá, J., García-Breijo, E., Ibáñez, J., Sogorb, T., Laguarda-Miro, N. and Garrigues, J. (2012) 'Low-cost capacitive humidity sensor for application within flexible RFID labels based on microcontroller systems', *IEEE Transactions on Instrumentation and Measurement*, **61** (2), 545–53.

Reverter, F., Jordana, J. and Pallàs-Areny, R. (2003) 'Internal trigger errors in microcontroller-based measurements', in *Proceedings XVII IMEKO World Congress*, Dubrovnik, Croatia, 22–27 June 2003, 655–8.

Reverter, F. and Pallàs-Areny, R. (2004) 'Effective number of resolution bits in direct sensor-to-microcontroller interfaces', *Measurement Science and Technology*, **15**, 2157–62.

Reverter, F. and Pallàs-Areny, R. (2005a) '*Direct sensor-to-microcontroller interface circuits*', *Design and Characterisation*, Barcelona (Spain), Marcombo.

Reverter, F., Jordana, J., Gasulla, M. and Pallàs-Areny, R. (2005b) 'Accuracy and resolution of direct resistive sensor-to-microcontroller interfaces', *Sensors and Actuators A*, **121**, 78–87.

Reverter, F. and Pallàs-Areny, R. (2006) 'Uncertainty reduction techniques in microcontroller-based time measurements', *Sensors and Actuators A*, **127**, 74–9.

Reverter, F., Gasulla, M. and Pallàs-Areny, R. (2007) 'Analysis of power-supply interference effects on direct sensor-to-microcontroller interfaces', *IEEE Transactions on Instrumentation and Measurement*, **56** (1), 171–7.

Reverter, F. and Casas, Ò. (2008a) 'Direct interface circuit for capacitive humidity sensors', *Sensors and Actuators A*, **143**, 315–22.

Reverter, F. and Casas, Ò. (2008b) 'Direct interface circuit for differential capacitive sensors', in *Proceedings IEEE I2MTC 2008*, Victoria, Canada, 12–15 May 2008, 1609–12.

Reverter, F. and Casas, Ò. (2009a) 'Interfacing differential resistive sensors to microcontrollers: a direct approach', *IEEE Transactions on Instrumentation and Measurement*, **58** (10), 3405–10.

Reverter, F., Horak, G., Bilas, V. and Gasulla, M. (2009b) 'Novel and low-cost temperature compensation technique for piezoresistive pressure sensors', in *Proceedings XIX IMEKO World Congress*, Lisbon, Portugal, 6–11 September 2009, 2084–7.

Reverter, F. and Casas, Ò. (2010a) 'Interfacing differential capacitive sensors to microcontrollers: a direct approach', *IEEE Transactions on Instrumentation and Measurement*, **59** (10), 2763–9.

Reverter, F. and Casas, Ò. (2010b) 'A microcontroller-based interface circuit for lossy capacitive sensors', *Measurement Science and Technology*, **21**(6), 065203 8.

Richey, R. (1997) *Resistance and capacitance meter using a PIC16C622,* AN611, Chandler, AZ: Microchip Technology Inc.

Sifuentes, E., Casas, Ò., Reverter, F. and Pallàs-Areny, R. (2008) 'Direct interface circuit to linearise resistive sensor bridges', *Sensors and Actuators A*, **147**, 210–15.

Sifuentes, E., Gasulla, M., Casas, Ò. and Reverter, F. (2009a) 'Improving the performance of direct interface circuits using the Vernier technique', in *Proceedings IEEE I2MTC*, Singapore, 5–7 May 2009, 1300–3.

Sifuentes, E. (2009b) *Sensor Autónomo Para Detectar Vehículos Estáticos*, PhD dissertation, Universitat Politècnica de Catalunya, Spain (in Spanish).

Sifuentes, E., Casas, Ò. and Pallàs-Areny, R. (2011) 'Wireless magnetic sensor node for vehicle detection with optical wake-up', *IEEE Sensors Journal*, **11** (8), 1669–76.

Stojanovic, R. and Karadaglic, D. (2007) 'A LED-LED-based photoplethysmography sensor', *Physiological Measurement*, **28**, N19–N27.

Vidal-Verdú, F., Oballe-Peinado, O., Sánchez-Durán, J.A., Castellanos-Ramos, J. and Navas-González, R. (2011) 'Three realizations and comparison of hardware for piezoresistive tactile sensors', *Sensors*, **11**, 3249–66.

Yurish, S.Y. (2009a) 'Universal resistance-to-digital converter', in *Proceedings CENICS 2009*, Sliema, Malta, 11–16 October 2009, 28–33.

Yurish, S.Y. (2009b) 'Universal capacitive sensors and transducers interface', in *Proceedings Eurosensors XXIII*, Lausanne, Switzerland, 6–9 September 2009, 441–4.

Yurish, S.Y. (2011) 'A simple and universal resistive-bridge sensors interface', *Sensors & Transducers Journal*, **10**, 46–59.

3

Capacitive sensors for displacement measurement in the sub-nanometer range

S. XIA, NXP Semiconductors, The Netherlands and
S. NIHTIANOV, Delft University of Technology,
The Netherlands

DOI: 10.1533/9780857099297.1.63

Abstract: The main challenge in designing an interface circuit for capacitive displacement sensors with sub-nanometer resolution is the large offset capacitance of the sensor. This offset capacitance is a result of the relatively large distance between the sensor electrodes, compared with the displacement range to be measured. From the different methods with which to interface capacitive sensors, a charge-balancing technique is used here to demonstrate a power-efficient way to remove the offset, while maintaining high resolution and short conversion time. As an example, the design of an oversampling capacitance-to-digital (CDC) converter based on a third order sigma/delta modulator is presented.

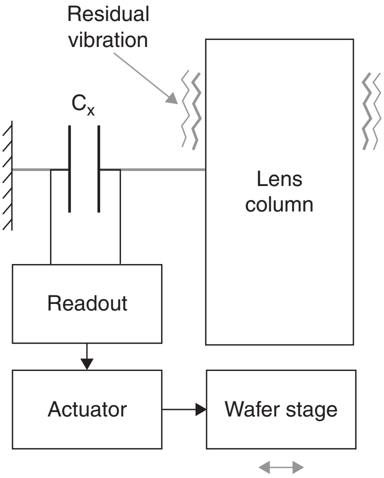
Key words: capacitive displacement sensor interface, switched-capacitor circuits, offset capacitance cancellation, incremental ΣΔ converter.

3.1 Introduction

Large offset capacitance often presents problems in capacitive sensor measurements. In this chapter, we discuss a capacitive sensor interface circuit, based on the charge-balancing technique, that removes the effect of offset capacitance electrically. High resolution is achieved with a relatively short conversion time.

3.2 Challenges for sub-nanometer displacement measurement with capacitive sensors

In precision mechatronic systems like wafer steppers, electron microscopes and so on, the position of critical mechanical components must be dynamically stabilized with sub-nanometer precision. This can be achieved with a servo loop consisting of a displacement sensor and an actuator, as shown in Fig. 3.1. For this application, capacitive displacement sensors offer a smaller



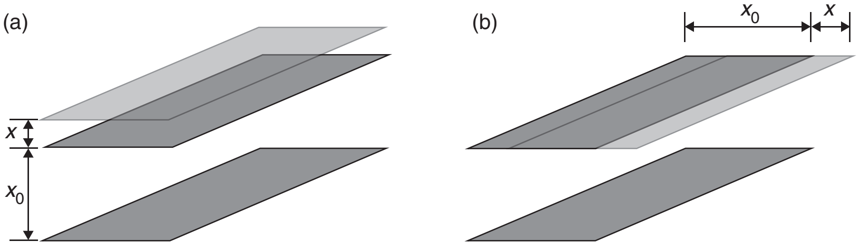
3.1 Example of a servo loop with a capacitive displacement sensor.

component size and lower cost, compared with other sensors, like optical interferometers, for example. Another important advantage of capacitive sensors is that they do not consume or generate electrical energy when converting displacement into electrical signal and, hence, they do not produce electronic noise, which is a prerequisite for obtaining extremely high resolution.

In Fig. 3.2, a parallel plate capacitive sensor is shown, the capacitance of which is a function of the gap and overlap between its two plates. The changes in both of these two physical parameters can be used to sense displacement. In both cases, the sensitivity to small displacement x is[1]:

$$\frac{dC}{dx} = \frac{C_0}{x_0}$$

where C_0 is the nominal capacitance at $x_{0,}$ and x_0 is the gap between the plates or overlapping width of the plate, respectively. It can be seen that decreasing x_0 can increase the sensitivity of the capacitive sensor. For typical geometries, the gap is much smaller than the overlap; thus, for high sensitivity capacitive displacement sensors, the variation in the plate gap is generally used. Although higher sensitivity comes at the cost of nonlinearity because, for the gap-closing transducer, the capacitance change is a known nonlinear function of displacement – if x_0 is known, this nonlinearity can be accounted for by backend processing.



3.2 Parallel capacitive displacement sensors working on changes of (a) gap closing, (b) overlapping area.

From the above discussion, it is clear that for maximum sensitivity, it is better that the gap between the plates is as small as possible. However, mechanical tolerances limit the minimum gap between the sensor electrodes to a few micrometers,[2] while in the targeted applications the required resolution has to be in the sub-nanometer range (typically, below 100 pm). At the same time, the displacement and/or the vibrations of the target to be measured are normally much less than 1 μm. In terms of capacitance values, if the nominal sensor capacitance at $x_0 = 10$ μm is 10 pF, the interface circuit needs to have an input-referred capacitance resolution of better than 100 aF. As a comparison, the variation of the sensor capacitance related to the displacement range is going to be much less than 1 pF, due to the low sensitivity of the capacitive sensor to displacement. So, the sensitivity of the displacement sensor becomes limited by the realizable gap distance. This also means that the readout circuit needs to have more than 17-bit resolution with respect to the nominal sensor capacitance. Furthermore, in servo-loop applications, often the measurement latency needs to be very low. The main challenge for the electronic interface is to provide high-resolution capacitance measurement with low measurement latency, which is challenging because of the trade-off between measurement time and resolution.

Since the nominal capacitance of the sensor is much greater than the sensor capacitance variation, it can be assumed that the sensor has a large offset capacitance. Luckily, there are ways to deal with this, as will be discussed in the next section.

3.3 Offset capacitance cancellation technique

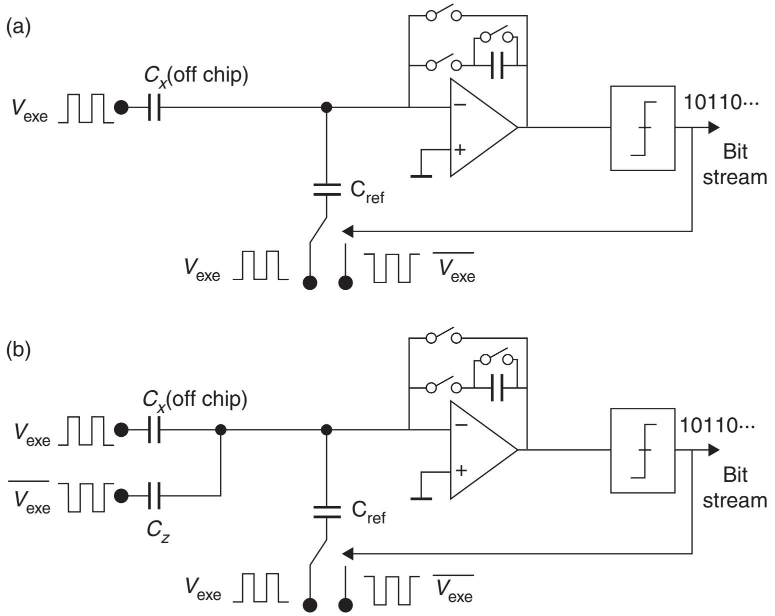
The existence of offset capacitance in a capacitive sensor often imposes a burden on the interface circuit. As the value of the sensor capacitance is dominated by the offset value, the interface circuit would have to resolve the relatively small capacitance variation on top of the offset capacitance. The offset capacitance will consume a large portion of the dynamic range of the interface, which would translate into the waste of system resources, such as power consumption and conversion time.

Fortunately, there are techniques to create circuits that cancel the effect of the offset capacitance.[3,4] Typically, a capacitive sensor interface would involve a virtual ground created by an operational amplifier (opamp) in order to make the readout of the capacitive sensor interface insensitive to the parasitic capacitances to ground at each of the two connections.[5,6] Such a virtual ground can also be used to cancel the effect of the offset capacitance, because a 'zoom-in' capacitor can be connected to the virtual ground and driven by an excitation signal that is opposite to that driving the sensor capacitor. The net charge that gets through the virtual ground and arrives at the feedback capacitor is then a function of the *difference* of the sensor capacitor and the zoom-in capacitor. By establishing the value of the zoom-in capacitor as equal to the offset capacitance of the sensor, the effect of the big sensor offset capacitance can be eliminated.

However, it should be noted that this function comes at a price. This is because the addition of the zoom-in capacitor increases the total capacitance at the input of the amplifier, which increases the noise gain of the amplifier.[7] Whether this penalty will be significant depends on the component values. For instance, if the value of the parasitic capacitance at the input of the amplifier (due, for example, to a long cable) is much greater than the sensor capacitance (and, hence, the zoom-in capacitance), then the penalty of introducing this zoom-in capacitance is negligible. In most cases, the benefit of using zoom-in capacitance over-rides the penalty introduced. However, it is still beneficial to bear this potential disadvantage in mind.

A typical way to measure capacitance is to compare it with a reference capacitance, by creating a ratio between them. There are many ways to obtain the capacitance ratio. However, the most popular method for obtaining high-resolution capacitive ratio is by using switched-capacitor circuitry combined with the charge-balancing principle.[8] This is because this method has a proven record of achieving both high resolution and high accuracy, at the expense of a generally slower conversion speed, because it basically trades resolution with time.[9] Figure 3.3a shows a block diagram of a capacitive sensor based on the charge-balancing principle. It is basically a switched-capacitor incremental ΣΔ converter.[10] In such a system, a virtual ground is also available to realize offset capacitance cancellation, as shown in Fig. 3.3b.

One of the most important advantages of switched-capacitor circuits in terms of realization is their relaxed requirement for an excitation signal. Mismatches in the excitation signals would normally translate into perceptions of capacitance difference, and would lead to an error in the readout circuit. For a switched-capacitor circuit, however, because the circuit is only sensitive to the final settled voltage values, the exact waveform of the two anti-phase excitation signals does not need to match perfectly, as long as



3.3 Block diagram of a capacitive sensor interface based on charge-balancing principle. (a) Illustrates the regular case, while (b) illustrates the case where a ‘zoom-in’ capacitor is added.

there is no charge loss and there is no signal-picking at the input of the convertor. This simplifies the realization a great deal.

The effective input of the interface becomes the difference between C_X and C_Z; that is, C_X–C_Z. When the value of C_Z is selected such that it equals the offset capacitance of C_X, the effect of the offset capacitance is completely removed. Then, a much lower C_{ref} can be applied that gives a measurement range that is just sufficiently broad to cover the small capacitance variation $\pm\Delta C_X$. In this way, the quantization requirement of the interface can be largely reduced. However, in practice it is difficult to ensure that C_Z fully cancels the offset capacitance of C_X, so some tolerances need to be kept in mind when designing the interface. In the next section, a design example will be given based on the method discussed above.

3.4 Capacitance-to-digital converter (CDC) with offset capacitance cancellation and calibration functions

In this section, a design example is presented that shows the benefits of using the zoom-in technique. The switched-capacitor incremental converter digitizes the capacitive ratio. The performance of the interface before and after using the zoom-in technique will be compared.

In this design example, the specifications are taken from a real application. The capacitive displacement sensor will have a nominal capacitance of 10 pF. The measurement range, in contrast, is less than 50 fF. The targeted resolution is better than 100 aF. The noise of the interface can be classified into thermal noise and quantization noise. Schreier *et al.*[7] provide an extensive overview on this topic. The total noise power is the sum of the thermal noise power and the quantization noise power.

Thermal noise in a switched capacitor circuit is directly linked to the size of the sampling capacitor.[7] For an incremental converter, which utilizes the oversampling technique, thermal noise can be further reduced by increasing the oversampling ratio. For a fixed conversion time, both methods for reducing thermal noise would require the interface to burn more power, either by driving larger capacitors or by using a faster clock. For a capacitive sensor interface circuit based on an incremental converter, the sensing capacitor acts as a sampling capacitor.[5] Its value is often fixed and cannot be changed arbitrarily by the designer. Therefore, the only way to reduce thermal noise in such an interface is by increasing the oversampling ratio.

Increasing the oversampling ratio reduces also the quantization noise level.[9] Depending on the loop filter structure, the slope at which the quantization noise is reduced will also change. Generally speaking, the higher the order of the loop filter, the faster the quantization noise reduces with the increasing oversampling ratio. Apart from using higher order loop filters, a multi-bit quantizer can be used, instead of a single-bit quantizer, in order to increase the speed of operation. Due to the noise-shaping effect on the quantization noise, the quantization noise decreases more rapidly than the thermal noise as the oversampling ratio is increased.

As discussed in the previous section, the zoom-in technique does not help to reduce the thermal noise level. So, basically, the thermal noise of the interface is determined to some extent by the absolute value of the sensor capacitor, which in our case is about 10 pF. Zoom-in, however, reduces the full-scale input range of the interface and therefore reduces the requirement on the quantization noise of the interface. So, it is important to point out that using the zoom-in technique only makes sense if, before it is applied, the quantization noise level is greater than the thermal noise level. In this example, the estimation of the thermal noise level is performed under the guidance provided by Schreier *et al.*[7] This shows that an oversampling ratio of about 100 is needed to reduce the input-referred thermal noise level to a sufficiently low level. So, in order to make the interface is not limited by quantization noise, we need to achieve a comparably low quantization noise level within an oversampling ratio of 100.

Without employing the zoom-in technique, the input signal range, including the offset capacitance, would be about 10 pF. Achieving capacitance resolution below 100 aF would require more than a 17-bit dynamic range. So, the question here is: how much resolution can be achieved with an oversampling ratio of 100, and will the corresponding input range be sufficiently large to cover the sensor capacitance variation during normal operation.

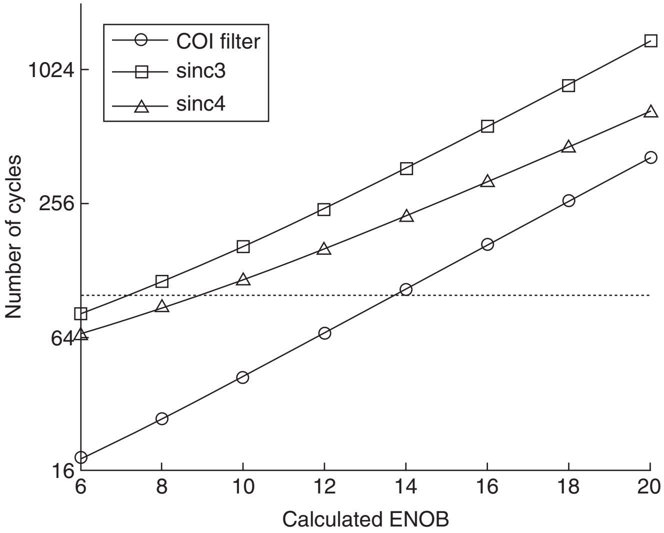
Although a multi-bit quantizer and a digital-to-analogue converter (DAC) can be used to reduce the required oversampling ratio, one of the most pronounced problems with this method is that the nonlinearity of the required multi-bit DAC will limit the linearity of the modulator. This is because the DAC is in the feedback and its error will be directly added to the error of the modulator. In order to solve this problem, dynamic element matching (DEM) has to be used.[9] However, this will lead to hardware complexity. A 1-bit DAC is intrinsically linear and thus does not suffer from this problem. Mainly, in order to reduce hardware complexity, a higher order loop filter is chosen as the optimum approach.

There is a theory, based on a linearized model, which can predict the necessary oversampling ratio for a particular resolution, with a certain loop filter order. However, at low oversampling ratio values, the prediction is often not very accurate, especially for incremental converters. A dedicated model was made that leads to complicated results for the prediction of the required oversampling ratio for an incremental converter.[11] The authors recommend using these results as guidelines only and rely on system-level simulation to obtain a better estimation. The conclusion based on the simulations results is that, with an oversampling ratio in the order of 100:

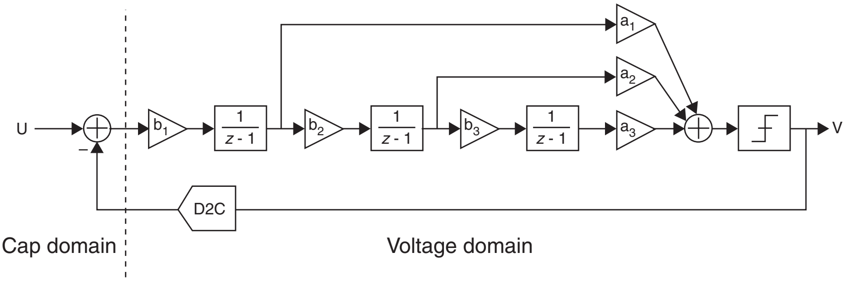
1. a second order loop filter will not meet the requirement of an input capacitance range of 50 fF, which is not acceptable for this application;
2. increasing the loop filter above third order brings negligible improvements. For this reason, a third order loop filter is chosen for this design.

With a third order loop filter, the input range is selected to be 200 fF, which brings sufficient margin to cover the variation of the sensor capacitance, while also providing some margin for the mismatch between sensor capacitance C_X and the offset-cancelling capacitance C_Z. This means that better than 12-bit resolution is needed when the zoom-in technique is applied.

With a third order loop filter, a very good quality decimation filter has to be used in the CDC in order to reach the required capacitance resolution within the limited conversion time. Figure 3.4 shows the achievable resolution of three different decimation filters for a $\Sigma\Delta$ converter with a third order loop filter, as a function of the number of clock cycles in one conversion, using the formulas given in Marcus (2005).[11] It can be seen



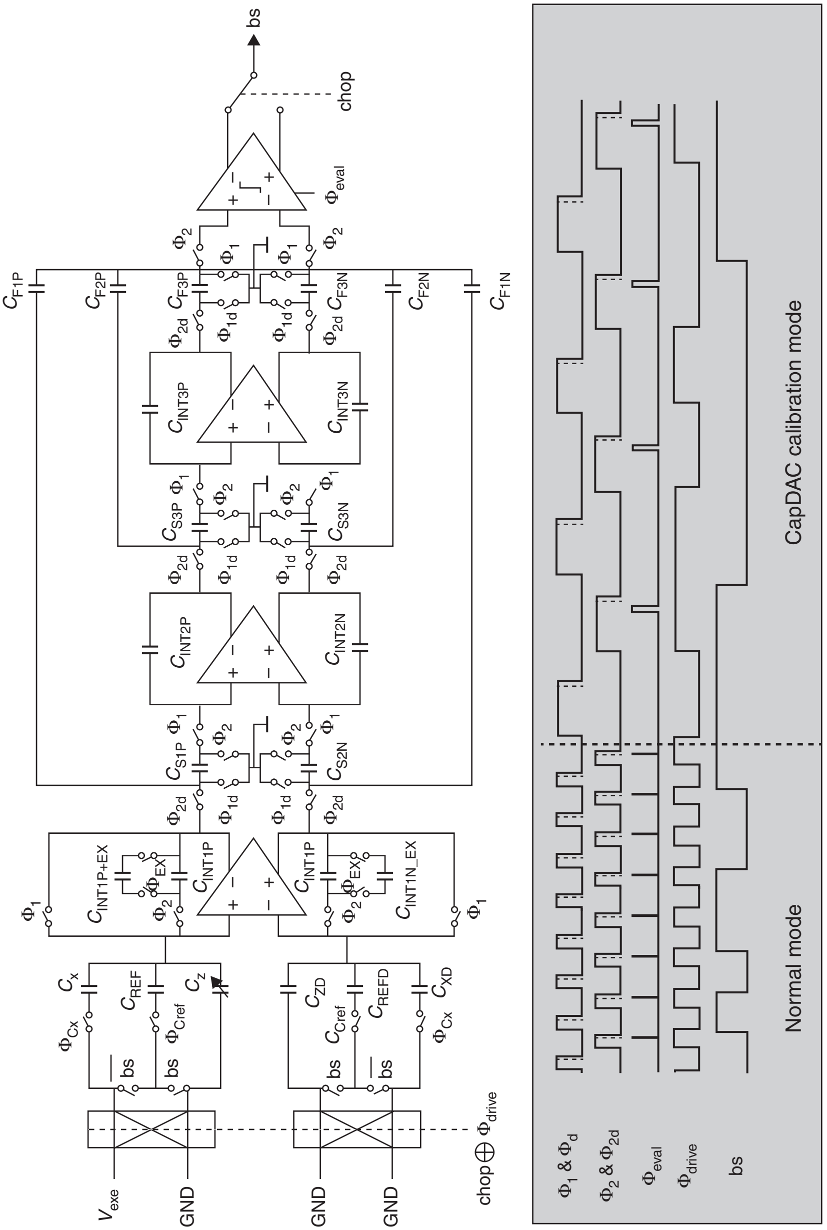
3.4 Achievable resolution of three different decimation filters for a ΣΔ converter with a third order loop filter as a function of the number of clock cycles.



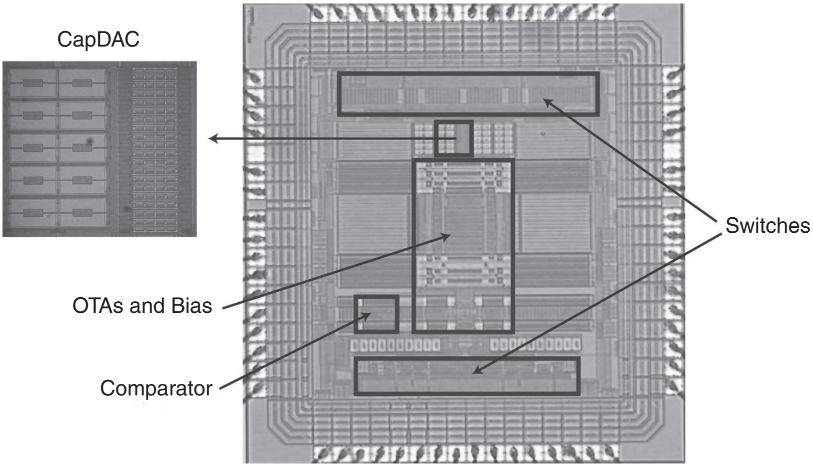
3.5 The chosen conventional feed-forward loop filter topology.

that only a cascade of integrators (CoI) decimation filter can reach better than 12-bit resolution within 100 clock cycles. The normally used sinc3 and sinc4 filters need 240 and 160 clock cycles, respectively, to reach the same resolution. In order to keep the conversion time low, a higher clock signal will be needed. For example, for a conversion time below 20 μs and 100 clock cycles, a 5 MHz clock signal is required.

As a comparison, if the zoom-in technique is not applied, the same third order incremental converter would need an oversampling ratio larger than 500 in order to reach a quantization noise level equivalent to 17 bits. So, in this example, by applying zoom-in, the conversion time with the same clock frequency will be five times faster. In addition to this, as will be discussed shortly,


V_{exe}
GND
GND
GND
chop ⊕ Φ_{drive}
$\overline{bs}$
bs
bs
$\overline{bs}$
Φ_{Cx}
Φ_{Cref}
C_x
C_{REF}
C_z
C_{ZD}
C_{Cref}
C_{REFD}
C_{XD}
Φ_{Cx}
Φ_1
$C_{INT1P+EX}$
Φ_{EX}
Φ_2
C_{INT1P}
C_{INT1P}
Φ_{EX}
Φ_2
C_{INT1N_EX}
Φ_1
Φ_{2d}
Φ_{1d}
C_{S1P}
C_{S2N}
C_{INT2P}
C_{INT2N}
C_{S3P}
C_{S3N}
C_{INT3P}
C_{INT3N}
C_{F3P}
C_{F3N}
C_{F1P}
C_{F2P}
C_{F2N}
C_{F1N}
Φ_{eval}
chop
bs
Φ_1 & Φ_d
Φ_2 & Φ_{2d}
Φ_{eval}
Φ_{drive}
bs
Normal mode
CapDAC calibration mode

3.6 Simplified circuit diagram of the interface, showing different timing diagrams in different operating modes.

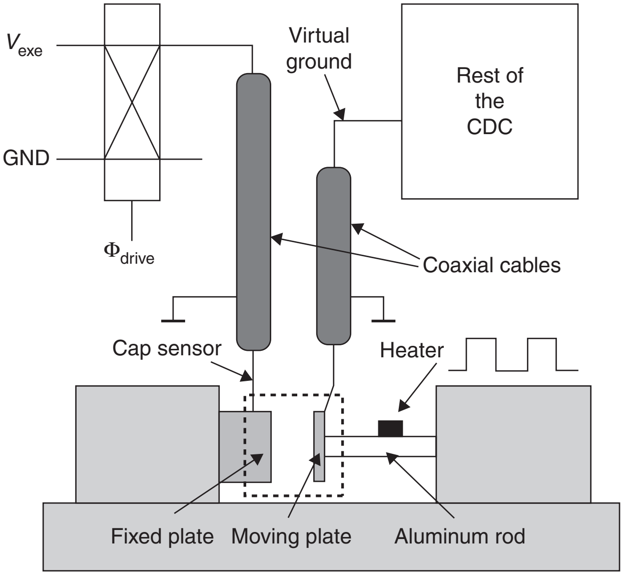


3.7 Die micrograph.

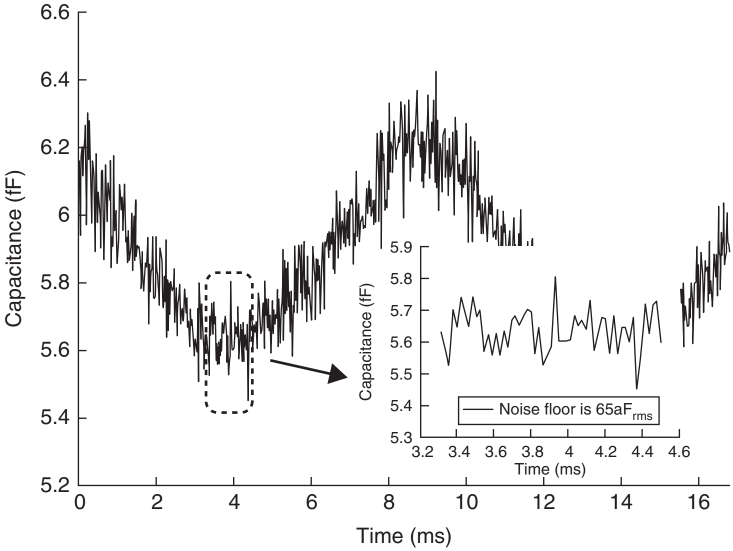
using the zoom-in principle also prevents slewing in the amplifier; this property can be further utilized to increase the clock frequency of the converter with the same level of current consumption of the amplifiers. In this sense, the actual benefit in applying the zoom-in principle is greater than five times.

Based on the above discussions, a capacitive sensor interface was designed, the block diagram of which is presented in Fig. 3.5. The interface is built around a switched-capacitor sigma-delta converter which is shown in Fig. 3.6.[12] The zoom-in capacitor C_Z is realized with a bank of on-chip capacitances CapDAC, to be able to easily adjust its value to match that of the sensor capacitance. CapDAC is basically a binary bank of capacitors that can be switched in and out by applying a digital code. The interface is fabricated in a standard 2P4M 0.35 μm complementary metal-oxide-semiconductor (CMOS) process and occupies an active area of 2.6 mm × 2 mm. A micro-photograph of the chip layout is shown in Fig. 3.7. It draws 4.5 mA from a 3.3 V supply, of which over 3.1 mA is consumed by the first operational transconductance amplifier (OTA) due to the need to drive large capacitances. The CapDAC and all the other on-chip capacitors are poly-poly capacitors because of their good linearity and temperature stability. For flexibility, in this design the control logic and the decimation filter were implemented off-chip.

To verify the performance of the CDC with an off-chip sensor, a test setup was built. As shown in Fig. 3.8, it consists of a fixed electrode in close proximity to an electrode attached to an aluminum rod. Heating the rod, *via* an attached resistor, changes its length and thus the capacitance between the two electrodes. The CDC is connected to the electrode on the rod *via* a short shielded cable, which adds extra parasitic capacitance (roughly 10


V_{exe}
GND
Φ_{drive}
Virtual ground
Rest of the CDC
Coaxial cables
Cap sensor
Heater
Fixed plate
Moving plate
Aluminum rod

3.8 Experimental set-up.



3.9 Measurements on a capacitive sensor that is mechanically actuated at 100 Hz.

pF) to ground. The small capacitance changes created by driving the heater with a 66 mHz, 8 mW square-wave are shown in Fig. 3.9. The actual value of the reference capacitor C_{ref}, as shown in Fig. 3.6, was found to be 91.8 fF by calibrating it against an external 10.3 pF surface-mount capacitor whose value had been measured by a HP4192A LF impedance analyzer (±100 fF inaccuracy). The capacitance changes were then found to correspond to 500 fFpp, while the noise on the measurements corresponded to 65 aFrms.

3.5 Conclusion

A practical way of removing the effect of the offset capacitance in a capacitive sensor has been presented. The realized interface circuit made use of the charge-balancing principle and achieved an effective capacitance resolution of 17 bits with a conversion time of 20 μs.

3.6 References

1. B. Boser, *Capacitive Interface Electronics for Sensing and Actuation*, Berkley: University of California, 21st Workshop on Advances in Analog Circuit Design, AACD, Valkenburg aan de Geul, The Netherlands (2012).
2. J.P. van Schieveen, J.W. Spronck, S. Nihtianov and R.H. Munnig Schmidt, ‘Integrated Auto Alignment and Calibration For High Resolution Capacitive Sensor System,’ Proc. EUSPEN, pp. 188–191 (2010).
3. J. Markus, J. Silva and G.C. Temes, AD7746 datasheet, Analog Devices, available online through http://www.analog.com.
4. D.Y. Shin, H. Lee and S. Kim, ‘A delta-sigma interface circuit for capacitive sensors with an automatically calibrated zero point,’ *IEEE Trans. Circuits and System II*, Vol. **58**, No. 2, pp. 90–94, February (2011).
5. J.O’Dowd, A. Callanan, G. Banarie and E. Company-Bosch, ‘Capacitive sensor interfacing using sigma-delta techniques,’ *Proc. IEEE Sensors*, October–November, pp. 951–954 (2005).
6. L.K. Baxter, *Capacitive Sensors: Design and Applications*, New York: John Wiley (1996).
7. R. Schreier, J. Silva, J. Steensgaard and G.C. Temes ‘Design-oriented estimation of thermal noise in switched-capacitor circuits,’ *IEEE Trans. on Circuits and Systems*, Vol. **52**, No. 11, pp. 2358–2368 (2005).
8. B. Wang, K, Tetsuya, T. Sun and G. Temes, ‘High-accuracy circuits for on-chip capacitive ratio testing and sensor readout,’ *IEEE Trans. Instrum. Measur.*, pp. 16–20, February (1998).
9. S.R. Norsworthy, R. Schreier and G.C. Temes (eds), *Delta-Sigma Data Converters: Theory, Design and Simulation*, Piscataway, New York: IEEE Press (1997).
10. J. Markus, J. Silva and G.C. Temes, ‘Theory and applications of incremental ΣΔ converters,’ *IEEE Trans. on Circuits and Systems*, Vol. **51**, No. 4, pp. 678–690 (2004).

11. J. Markus, PhD Thesis, Budapest University of Technology and Economics, Department of Measurement and Information Systems, Budapest, Hungary (2005).
12. S. Xia, K. Makinwa and S. Nihtianov, ‘A Capacitance-to-Digital Converter for Displacement Sensing with 17b Resolution and 20 μs Conversion Time,’ Digest of technical papers, ISSCC, 2012.

4

Integrated inductive displacement sensors for harsh industrial environments

M. R. NABAVI, Catena Microelectronics BV,
The Netherlands and S. NIHTIANOV,
Delft University of Technology, The Netherlands

DOI: 10.1533/9780857099297.1.76

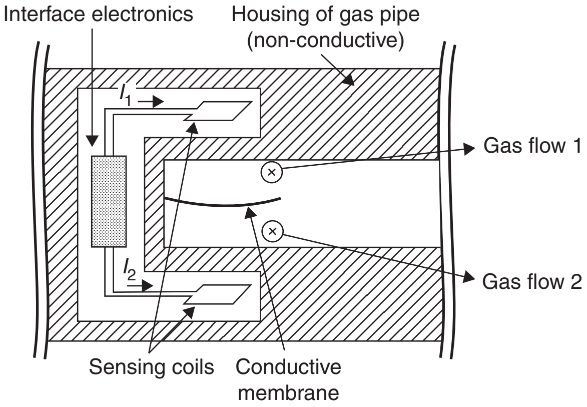
Abstract: This chapter discusses eddy-current displacement sensors (ECSs) – one of the most widely-used measurement techniques in harsh environments. First, in Section 4.1, we briefly discuss ECS applications in industry. In Section 4.2, the operation principle, the limitations and important design considerations of this type of sensor are surveyed. The ECS performance limitations are addressed in Section 4.3, considering the most demanding industrial applications. Special attention is given to the excitation frequency, the design of the sensor and its electronic interface. Sections 4.4 and 4.5 provide a review of state-of-the-art ECS interfaces. The chapter ends with conclusions and design perspectives for ECS systems.

Key words: eddy-current sensor, displacement measurement, integrated interface, harsh environment, stability, resolution.

4.1 Introduction

Displacement sensors are extensively employed in industry for measuring the position and movement of objects, as well as for measuring other physical quantities which can be converted into movement, such as pressure, acceleration and so on. Depending on the industrial application, the preferred types of non-contact displacement sensors are optical/laser, capacitive and inductive (eddy-current) sensors (ECSs).[1,2] Incremental optical sensors, such as interferometers and encoders, provide the best combination of accuracy and measurement range, but they are bulky and quite expensive due to their sophisticated structure. Capacitive sensors and ECSs[1] are preferred when the absolute position of a target needs to be measured with a limited dynamic range. These sensors are compact, robust, stable, accurate and relatively low cost.

A closer look at the principle of operation of capacitive sensors and ECSs reveals some similarities. Both sensors represent reactance in electrical



4.1 A simplified differential configuration of ECS sensors for measuring gas pressure.

circuits, which means that, they do not consume any electrical energy, and for this reason they are practically noiseless. The structure of an ECS with a flat sensing coil is similar to the structure of a capacitive sensor. Both sensors, as well as their targets, can be built using similar materials. Therefore, the difference in performance of these two sensor types can be mostly related to their sensitivity to working environment variations, the way of interaction with their targets, and the performance of the utilized electronic interfaces.

ECSs have the potential for very precise position/displacement measurement in the micrometer and even in the nanometer range.* Unlike capacitive sensors, ECSs are highly insensitive to variations in the working environment such as pressure, humidity, contamination. Furthermore, these sensors do not require an electric contact with the target. This is an important advantage in many applications where no contact with the target is allowed (such as measuring the position of a magnetically levitated target[3]).

Figure 4.1 shows an application example in which the difference between the gas pressure in Gas flow 1 and Gas flow 2 is the primary measurand.[4] The pressure difference deflects the conductive membrane positioned between the two gas flows. The target deflection causes the inductance of one sensor to increase and that of the other to decrease. Such a differential set-up helps increase the sensitivity to the target movements compared with the effectiveness of employing only one sensor. An important advantage of employing an ECS is its immunity to the dielectric properties of the flowing gas.

---

* Eddy currents are induced electrical currents that flow in a circular path. Their name is derived from the 'eddies' that are formed when a liquid flows in a circular path around obstacles.

To exploit the above-mentioned advantages of ECSs fully and to create a high-performance displacement sensor based on the eddy-current principle, their operation principle as well as their short-comings have to be taken into account. In the following section, we point out these aspects.

4.2 Principles of operation and practical limitations for eddy-current sensors (ECSs)

In this Section the operation principle of eddy-current displacement sensors is investigated. Based on the operation principle, practical limitations of ECSs are analyzed.

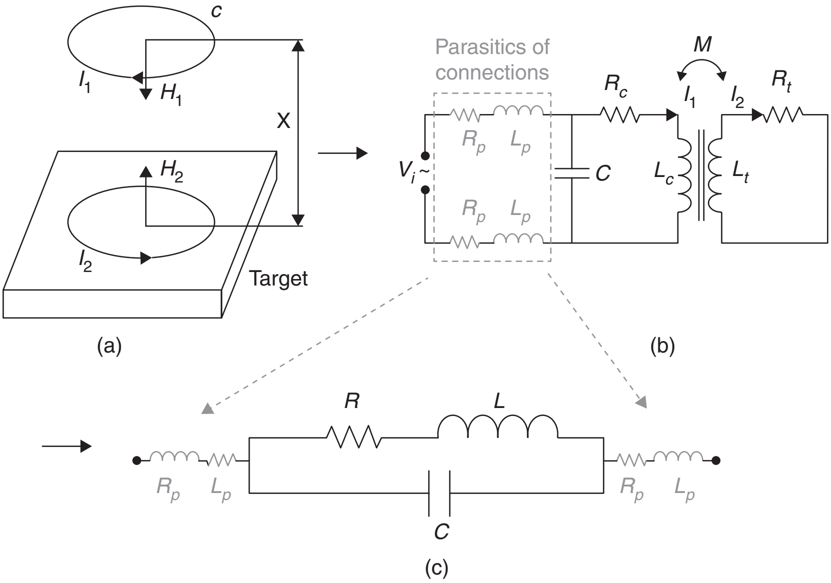
4.2.1 Sensor operation principles

An ECS measurement system includes a sensing coil set in front of a metallic (conductive) target. Exciting the coil with an AC current generates an electromagnetic field. The alternating magnetic field applies Lorentz's force to the electrons in the target. This results in a flow of electrons (i.e., circulating eddy currents) in the target, in a plane perpendicular to the direction of the magnetic field. Based on Lenz's law, the eddy currents generate a secondary magnetic field that opposes the primary excitation field, as shown in Fig. 4.2a. The closer the target is to the coil, the higher is the density of the eddy currents. In other words, there is a magnetic field interaction between the coil and the target. This interaction can be electrically modeled using an air-core transformer in which the mutual inductance between the primary and secondary sides depends on the sensor-target distance X, as shown in Fig. 4.2b. Since the eddy currents dissipate electrical power, a parasitic series resistance is usually considered in their electrical model.

On the basis of the equivalent circuit, a simplified electrical model of the sensor is shown in Fig. 4.2c. In the model, C represents the inter-winding parasitic capacitance of the sensor, while R_p, L_p (indicated in grey in Figs 4.2b and 4.2c) represent the parasitic contribution of the sensor connections. Expressions [4.1] and [4.2] define the equivalent inductance and series resistance of the sensor as[5]:

$$L = L_c - \frac{(2\pi f_{\text{exc}})^2 \cdot M^2}{R_t^2 + (2\pi f_{\text{exc}} L_t)^2} L_t \qquad [4.1]$$

$$R = R_c + \frac{(2\pi f_{\text{exc}})^2 \cdot M^2}{R_t^2 + (2\pi f_{\text{exc}} L_t)^2} R_t, \qquad [4.2]$$



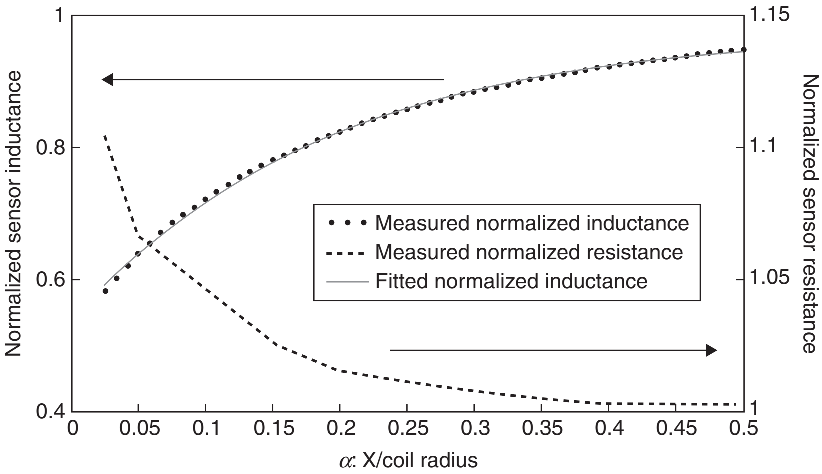
4.2 (a) Operating principle of an ECS; (b) ECS model including an air-core transformer; (c) simplified electrical model of an ECS (R and C are parasitic components which should be minimized when designing the coil).

where L_c and R_c are the inductance and series resistance of the coil without a target, L_t and R_t are the inductance and series resistance of the eddy current loop, f_{exc} is the frequency of the excitation signal and M is the mutual inductance between the coil and the eddy current loop (see Fig. 4.2).

Attempts to achieve an analytical formulation of the ECS inductance-displacement transfer function have been reported.[5–8] However, since many parameters contribute to such analyses, the mathematics become very complicated, which makes obtaining a perception of the physical behavior of the sensor too difficult.[8] Figure 4.3 shows the empirically obtained dependence between the normalized inductance and parasitic series resistance (of a prototype with a flat sensing coil) and the ratio α between the sensor-target distance X and the coil radius. The following exponential function is fitted to the measured data (Fig. 4.3):

$$\frac{L(\alpha)}{L_c} \approx a - b \cdot e^{-c \cdot \alpha} \qquad [4.3]$$

where L_c is the coil inductance without the target, and a, b and c are curve-fitting coefficients. The value of a is close to unity, which is in agreement



4.3 Measured inductance and series resistance (normalized by L_c and R_c) of a flat sensor coil for different distances X from a copper target (the right vertical axis corresponds to the dashed line curve); the fitted normalized inductance curve, shown in a solid line, is obtained from Expression [4.3].

with the fact that, when the target is far away (i.e., $\alpha \to \infty$), then the sensor inductance tends to its intrinsic value (i.e., $L(\alpha) \to L_c$). The parameters b and c relate to the sensor and target parameters such as shape, geometry and material.** The maximum deviation between the exponential fitted curve and the measured values does not exceed a few percent. When the target moves further away, the parasitic series resistance of the sensor decays slightly. This can be explained by Lenz's law: moving the target away implies a smaller secondary magnetic field and, hence, a smaller amount of eddy current, as well as less dissipated energy (i.e., the equivalent sensor resistance lowers). Typically, the changes in the series resistance are negligible compared with the sensor inductance variations, as the target moves. It should be mentioned that knowing the sensor transfer function nature facilitates the development of efficient empirical linearization algorithms/functions for ECS measurement systems aids in the design process.

The sensor operation principle predicts a few performance features which can potentially deteriorate its overall performance; the most important are discussed in the following section.

** One may investigate the general relation between these parameters and the curve-fitting coefficients. In a case study, it was discovered that: $a = 0.97, b = 0.44$ and $c = 5.38$ using a flat sensing coil with a radius of 6 mm and $L_c \approx 38.5$ nH (and series resistance of 40 mΩ).[12]

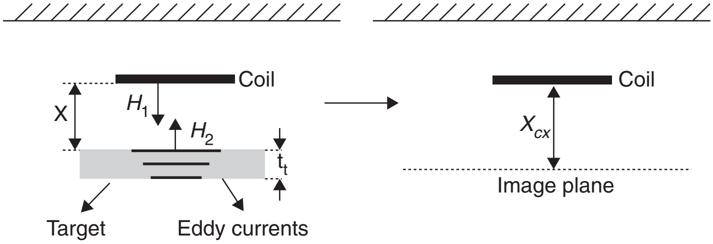
4.2.2 Limitations of eddy-current sensors

Skin effect

The intensity of the eddy currents weakens as they penetrate deeper into the target. This phenomenon is called the 'skin effect'. The penetration (skin) depth δ of the eddy currents is defined as:

$$\delta = \sqrt{\frac{2}{\omega_{\text{exc}}\mu\sigma}} \tag{4.4}$$

where μ is the target permeability, σ is its conductivity and $\omega_{\text{exc}} = 2\pi f_{\text{exc}}$. The standard penetration depth is determined by the depth at which the eddy current intensity reaches approximately 37% (i.e., $1/e$) of its intensity at the target surface (see also Fig. 4.4). Table 4.1 shows the skin depth for different frequencies at room temperature for some materials typically employed in industrial applications. If δ exceeds the thickness of the target being tested, the sensitivity of the sensor drops because a considerable part of the eddy currents is lost.[9]



4.4 Replacing the target with a penetrating eddy current by a lossless image plane.

Table 4.1 Penetration depth for different materials and excitation frequencies

Metal	Conductivity ($\times 10^6$ S/m)	Permeability ($\times 10^{-6}$ H/m)	Penetration depth (μm)					
			10 kHz	100 kHz	1 MHz	10 MHz	100 MHz	1 GHz
Copper	58.00	1.25	660.0	210.0	66.0	21.0	6.6	2.0
Aluminium	38.00	1.25	820.0	260.0	82.0	26.0	8.2	2.6
Stainless steel	1.30	1.25	4378.0	1392.0	437.0	140.0	44.0	14.0
Titanium alloy	0.59	1.25	6600.0	2100.0	660.0	210.0	66.0	21.0

Moreover, a penetration depth comparable to the thickness of the target will degrade the measurement reliability. Using the complex image method, a target with eddy currents penetrating into it can be replaced by a superconductor image plane at a complex distance of X_{cx}, as shown in Fig. 4.4, so that the equivalent magnetic potential (related to $\overrightarrow{H_1} + \overrightarrow{H_2}$) remains the same in both cases. The parameter X_{cx} is obtained from[10,11]:

$$X_{cx} = X + \frac{1-j}{2}\delta \cdot \tanh\left[(1+j) \cdot \frac{t_t}{\delta}\right] \tag{4.5}$$

where δ is the penetration depth and t_t is the target thickness, as shown in Fig. 4.4. Although the complex distance X_{cx} may not have a physical interpretation, it helps to understand the behavior of the sensor.[12] The second term in Equation [4.5] indicates the effect of δ and t_t on the sensor performance. Knowing this allows the sensor cross-sensitivities to the target thickness and the penetration depth to be approximated by:

$$\frac{\partial X_{cx}}{\partial t_t} = 1 - \tanh^2\left[(1+j) \cdot \frac{t_t}{\delta}\right] \tag{4.6}$$

and

$$\frac{\partial X_{cx}}{\partial \delta} = -\frac{t_t}{\delta} \cdot \left\{1 - \tanh^2\left[(1+j) \cdot \frac{t_t}{\delta}\right]\right\} + \frac{1-j}{2} \cdot \tanh\left[(1+j) \cdot \frac{t_t}{\delta}\right] \tag{4.7}$$

With regard to Expression [4.5], the cross-sensitivity to the target thickness can be ignored when $\tanh\left[(1+j) \cdot (t_t/\delta)\right] \to 1$, which is obtained for values of δ smaller than $t_t/3$. In practical applications, satisfying this condition helps to eliminate the measurement cross-sensitivity to a probable intervening conductive object behind the main target. It should be mentioned that this design choice also causes the eddy currents generated in the target to be maximized, which improves the sensitivity of the sensor to the position of the target X.

It is important to mention that the cross-sensitivity to the target electrical parameters also has to be taken into account in precision measurements. As an example, the cross-sensitivity to the target temperature is given by (using Expression [4.7]):

$$\left|\frac{\partial X_{cx}}{\partial T}\right| = \left|\frac{\partial X_{cx}}{\partial \delta} \cdot \frac{\partial \delta}{\partial \rho} \cdot \frac{\partial \rho}{\partial T}\right| = \frac{1}{2\sqrt{2}} \cdot \gamma \cdot \delta \tag{4.8}$$

where ρ is the target resistivity and γ is the temperature coefficient of the target resistivity. Expression [4.8] presents a fundamental maximum design value for δ which keeps the resulting temperature drifts within a desired stability budget (provided that the electronic interface is ideally thermally stable).

Parasitic effects

In addition to the sensing coil main parameter, which is the inductance, inevitable parasitic components – such as the coil intrinsic series resistance, the inter-winding capacitance and the parasitics originating from the electrical connection between the sensor and the interface electronics – contribute to the measurement result. However, in most practical cases the parasitics can be tolerated if they do not deteriorate the measurement resolution and/or stability. This implies that, first, they have to be minimized and, second, their variations (e.g., with temperature) have to be within the design error budget.

Due to the capacitive coupling between the windings of the coil, an inevitable inter-winding parasitic capacitance appears in the sensor model (see Fig. 4.2c). It should be mentioned that structures such as multi-layer coils or densely wound coils introduce an increased parasitic capacitance value. With the increase of the sensor excitation frequency, the parasitic capacitor can short-circuit the sensor inductance, which causes the sensor sensitivity to drop. The excitation frequency requirements lead to important design considerations of the sensor (and also of the interface), particularly in precision applications. These considerations are further discussed in Section 4.3.

Limited sensing coil quality factor

As shown in Figs 4.2c and 4.3, the sensor suffers from a parasitic series resistance.[†] This resistance limits the sensor quality factor Q and, hence, the filtering effect against external interferences. Regarding Expressions [4.1] and [4.2], the quality factor Q can be expressed by:

$$Q = \frac{L\omega_{\mathbf{exc}}}{R} = \frac{\left(L_c - \left(\left(\left((2\pi f_{\mathbf{exc}})^2 \cdot M^2\right)/R_t^2 + (2\pi f_{\mathbf{exc}} L_t)^2\right) L_t\right)\right).(2\pi f_{\mathbf{exc}})}{R_c + \left(\left((2\pi f_{\mathbf{exc}})^2 \cdot M^2\right)/\left(R_t^2 + (2\pi f_{\mathbf{exc}} L_t)^2\right)\right) R_t} \qquad [4.9]$$

[†] In the case of using targets with high enough conductivity (e.g., Copper or Aluminium) the coil parasitic resistance R_c becomes quite dominant in Equation [4.2].

A sufficiently high excitation frequency improves the quality factor of the sensor and helps in detecting a small inductance variation in the presence of a parasitic series resistance.[‡]

Frequency dependence

Expressions [4.1] and [4.2] indicate the frequency dependence of the equivalent inductance and resistance of the sensing coil. This effect introduces a cross-sensitivity to a variation of the excitation frequency. However, provided that $f_{\text{exc}} \rightarrow \infty$, we can write[5]:

$$L \rightarrow L_c - \frac{M^2}{L_t} \qquad [4.10]$$

Expression [4.10] shows that increasing f_{exc} helps to lower the unwanted cross-sensitivity to the excitation frequency variations.

4.3 Design requirements in precision industrial applications

According to the above-mentioned analyses, increasing the excitation frequency of the ECS can play an important role in advanced industrial applications. Examples for such applications are: measuring the position of a thin conductive membrane, or operation in a harsh environment with a limited temperature control. An example of design requirements, dictated by these kinds of applications, are given next.

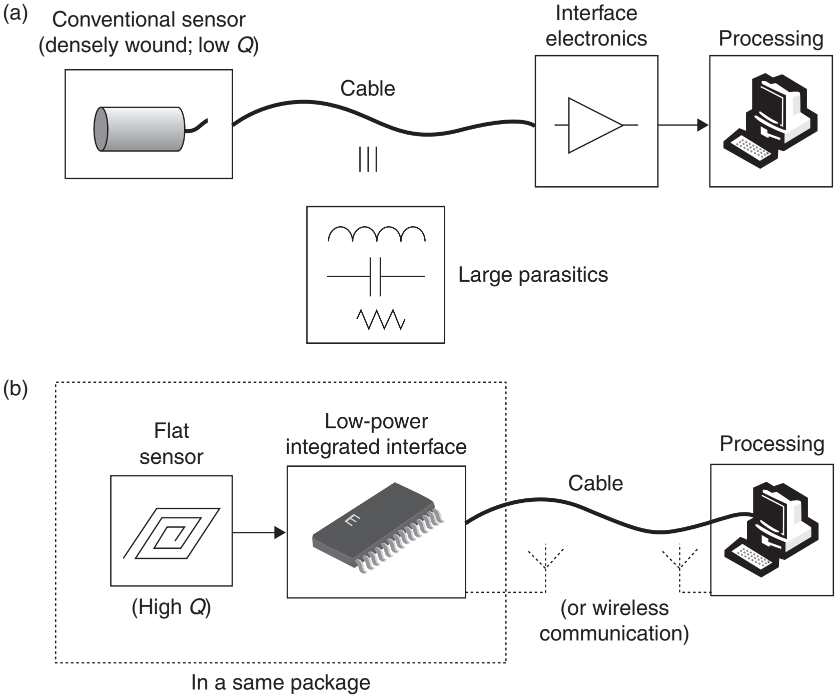
1. With respect to the design criterion $\delta \leq t_t/3$, the use of a 100-μm thick copper/aluminium target results in an excitation frequency of ~10 MHz.
2. Based on Expression [4.8], a thermal instability budget of the ECS of ×10 nm/°C (using a copper/aluminium target) requires ~20 MHz excitation frequency, which is more than one order of magnitude higher than the maximum popular values (of a few MHz).[§]

[‡] Note that for a given material, increasing the excitation frequency causes the parasitic resistance in the path of the eddy currents to increase due to the lowering of δ. However, this increase relates to $\sqrt{f_{\text{exc}}}$ while the inductive impedance increase relates to f_{exc}. Therefore, in general, an increase in f_{exc} helps to detect the inductive impedance against the parasitic series resistance.

[§] Performing a temperature measurement in order to compensate for the thermal drift is a possible alternative option. However, in many displacement measurement systems, in which there is no easy access to the target, measuring the target temperature can be a considerable challenge.

However, increasing the excitation frequency when using a typical densely wound coil is problematic since its relatively large parasitic capacitance (see Figs 4.2c and 4.5a) shorts the inductance of the sensor as f_{exc} increases. Flat coils with fewer turns and sufficient distance between their turns are preferred since their associated parasitic capacitance is quite limited. Moreover, such coils with a flat structure are volume-efficient and have an improved mechanical stability. However, for a given sensor area this kind of sensor offers a very limited inductance (e.g., ×10 nH). To avoid the relatively large parasitic inductance of a connection cable, which will appear in series of the small sensor inductance, integrating the sensor and the interface, or assembling them in one package, is indispensable (see Figs 4.2b, 4.2c and 4.5b).

It is important to point out that increasing f_{exc} and utilizing a flat coil with a very small series resistance R (e.g., ×10 mΩ) considerably enhances the sensor quality factor ($Q \propto f_{\text{exc}}, R^{-1}$). Improving quality factor Q, together with removing the connection cables and integrating the sensor and the


(a)
Conventional sensor
(densely wound; low Q)
Cable
Interface
electronics
Processing
Large parasitics
(b)
Flat
sensor
(High Q)
Low-power
integrated interface
Cable
Processing
(or wireless
communication)
In a same package

4.5 An ECS system with functional blocks: (a) conventional realization (not compatible with high excitation frequency required in advanced applications); (b) state-of-the-art implementation with a low-power integrated interface (suitable for high-frequency and high-precision operation).

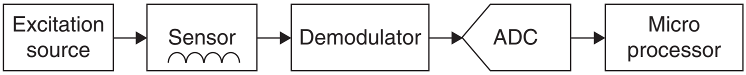
interface, improves the measurement immunity to stray magnetic fields (Fig. 4.5b). This is particularly important in harsh industrial environments.

Integrating the sensor and the interface, together with applying an increased f_{exc}, promises a considerable performance enhancement at system level; however, it also introduces challenges to the interface electronics. Operating with a higher excitation frequency implies an increase in the power consumption of the interface, which is needed to achieve a given resolution, stability and bandwidth. High power consumption is associated with increased heat generation, causing an increase in the sensor temperature, and, hence, thermal drift of the output signal. Furthermore, introducing a local heat source is unacceptable in many working environments. Due to the fundamental trade-offs in the integrated circuits technology between power consumption, operating frequency, stability and resolution, achieving a high performance at a higher f_{exc} becomes challenging. In Sections 4.4 the latest advances in ECS interfaces are discussed, in order to conceptualize the pros and cons of the existing ECS design concepts.

4.4 State-of-the-art ECS interfaces

Figure 4.6 shows the standard architecture of an ECS interface. The first stage excites the sensor, and the sensor inductance (as a function of the target position) modulates the excitation signal. Most often, a frequency modulation (FM) or an amplitude modulation (AM) is used.[13] A front-end stage (e.g., voltage or current amplifier) conditions the information-carrying signal before supplying it to the following stages. The next stage demodulates the displacement information to the baseband (BB). Eventually, an analog-to-digital converter (ADC) digitizes the displacement information and applies it to a micro-processor for data processing (or, in some applications, to a radio interface for data transmission). It is noteworthy that, depending on the implementation method, a number of the functions may be combined in a single circuit.

Since the sensor impedance can be very limited (e.g., a few ohms at 20 MHz excitation frequency), employing conventional impedance measurement approaches – such as two-port measurement – is not feasible. More specifically, such limited impedance generates a very small signal at a



4.6 Standard functional diagram of an ECS interface.

relatively high frequency. The result of this is increased power consumption of the conventional signal conditioning solutions.

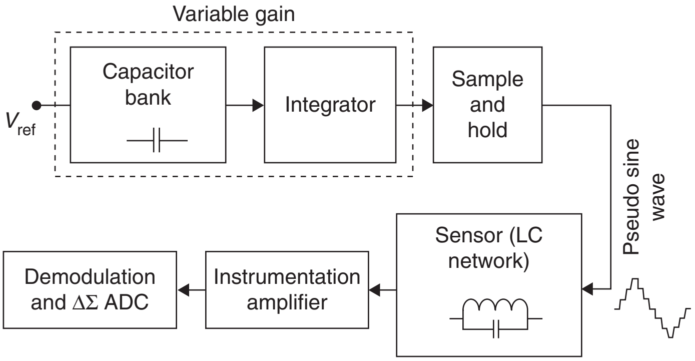
A more power-efficient alternative to interface an ECS is to use the sensing coil as part of a resonator or an oscillator.[3,14–16] A number of recently reported ECS interfaces employ this approach. They are evaluated in the next subsections. The leading assessment criterion is the potential of the interface to use a high sensor excitation frequency (in the ×10 MHz range). This undoubtedly relates to the method of exciting the sensor, which is a critical design decision in any ECS system.[17]

4.4.1 Utilizing external switched-capacitor oscillator and LC resonator

Oberle[3] proposes a power-efficient ECS interface (for a differential 2D displacement measurement system) which operates at ×100 kHz excitation frequency. To excite the coils, a switched-capacitor (SC) oscillator is employed to generate an approximation of a sine wave of 312.5 kHz (a pseudo-sine wave), as shown in Fig. 4.7.

The levels of the excitation signal are determined by V_{ref} and the variable gain of the SC integrator. To generate the 312.5 kHz excitation frequency, which is slightly below the minimum expected resonant frequency of the sensor (circuit) network, the switching operation is performed at a 5 MHz rate. This signal generator consumes more than 4 mW. It also uses an auto-zero (AZ) mode of operation to suppress the effect of its amplifier offset and, hence, the drift of the generated excitation signal.[18] The pseudo-sine wave is applied to the sensor network, which also acts as a bandpass filter (suppressing the excitation signal harmonics). The signal amplitude is modulated by the sensor impedance and amplified using the instrumentation amplifier stage. In order to demodulate and digitize the displacement information, the sensor signal is applied to a following mixer and a ΔΣ ADC. A dynamic range of 10 bits and a signal bandwidth of 10 kHz, with a low-power dissipation of 10 mW, are reported as using this interface design method.

However, extrapolating this to a 20 MHz excitation frequency, the SC oscillator would require a switching frequency of above 350 MHz. Such a high frequency leads to a higher power consumption and may cause undesired cross talk (e.g., between the channels of a multi-sensor interface or differential set-ups, like the one shown in Fig. 4.1). Moreover, implementing so many active blocks operating at high frequency results in significant increase in power dissipation, which is not acceptable in power-efficient applications.



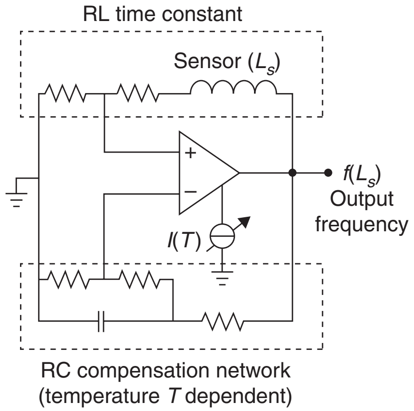
4.7 Functional block diagram of the interface using an SC oscillator for sensor excitation (with f_{exc} in ×100 kHz range).[3]

4.4.2 Relaxation-oscillator based interface

A high excitation frequency ECS interface is reported using the self-oscillation principle.[9] In this interface, as illustrated in Fig. 4.8, a relaxation oscillator is employed, the output frequency of which is modulated by the target displacement. The time constant of the positive feedback network depends on the sensor inductance and, therefore, the inductance variations modulate the oscillation frequency. It is reported that the excitation frequency varies from 10 to 30 MHz, depending on the target position. In this design, due to the sensitivity of the time constant to the temperature, a compensation RC network in the negative feedback path of the amplifier is employed, together with a temperature-controlled biasing current. The output frequency is measured by a counter to determine the sensor inductance. An advantage of this approach is its simplicity, while it achieves a good dynamic range of more than 12 bits.

Although the excitation frequency is within the desired range, the reported power consumption of 100 mW is too high for many applications. One reason for using so much power is the lack of an LC resonator network and, hence, the lack of boosting of the low sensor impedance (e.g., in few tens of Ω range). Therefore, a considerable amount of current is required to drive a sensor with such a low impedance at a relatively high excitation frequency.

It is necessary to mention that FM modulation is not the preferred choice for an ECS system. When using this modulation principle, most often for reasons related to the desired dynamic range, the interface has to be designed in such a way that the frequency varies within a relatively wide range as the target moves (±50% frequency change).[19] Such a wide frequency variation significantly changes the penetration depth of the eddy currents in the target.[20] Therefore, an additional nonlinearity in the sensor transfer function



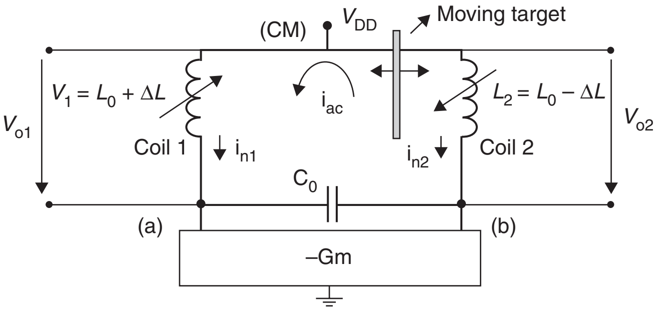
4.8 Interface implemented in Kejik *et al.* [19]

will appear due to the non-linear relation between f_{exc} and δ (X_{cx}) (see also Expression [4.5]). AM modulation with limited frequency variation is the preferred solution to avoid further deterioration of the sensor linearity.

4.5 ECS interfaces with an LC oscillator and ratiometric measurement

A front-end stage including an LC oscillator (depicted in Fig. 4.9) has been proposed.[21] The stage utilizes two differentially connected sensors which are excited by the oscillator signal at 20 MHz $\left(2\pi f_{\text{osc}} = 2\pi f_{\text{exc}} \simeq 1/\sqrt{(L_1 + L_2)\cdot C_0}\right)$. It is also possible to utilize only one sensing coil, while the other coil is used as a reference, but then a similar nonlinearity problem to the previously discussed interface in Section 4.4.2 occurs. The sensor inductances in differential mode modulate mainly the amplitude of the oscillator output signals. Although two inductors with basically different values are utilized, only one oscillation frequency appears since in this front-end stage only a single resonator is formed.

An important feature of this stage is that the noise levels at both outputs are correlated. This property can be utilized for the purposes of the noise suppression operation. Since, in the vicinity of the oscillation frequency, the AC current is trapped in the resonator, both coils experience the same current i_{ac}.[4] This implies that i_{ac}, together with the noise and drift associated with this current, modulates the amplitude of both output signals in an identical multiplicative way, which results in a common-mode multiplicative error. A similar conclusion can be drawn for the temperature dependence of the sensing coils[4] – in a first order analysis, similar thermal expansion coefficients are added to the values of both sensing inductances. Due to the



4.9 LC front-end oscillator with a single resonator (including two sensors) and correlated noise levels at the outputs ($i_{n1} = i_{n2}$ at the vicinity of oscillation frequency; in differential sensor mode, the variation of f_{osc} is very small since $L_1 + L_2 \approx cte$).

multiplicative nature of the disturbance factors, utilizing a ratiometric measurement can significantly enhance the measurement resolution and stability of performance. The ratiometric function is defined as:

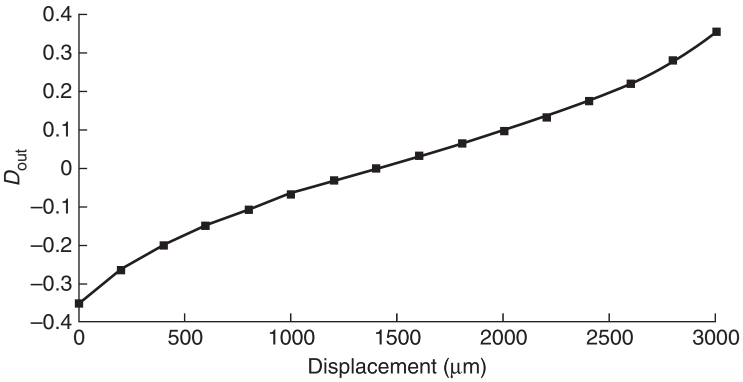
$$D_{\text{out}} = \frac{|V_{O1}| - |V_{O2}|}{|V_{O1}| + |V_{O2}|} \simeq \frac{|L_1 \omega_{\text{osc}} \cdot I_{\text{ac}}| - |L_2 \omega_{\text{osc}} \cdot I_{\text{ac}}|}{|L_1 \omega_{\text{osc}} \cdot I_{\text{ac}}| + |L_2 \omega_{\text{osc}} \cdot I_{\text{ac}}|} = \frac{\Delta L}{L_0} \qquad [4.11]$$

It is notable that, by using this readout function, the linearity of the differential system is preserved and the transfer characteristic looks like very similar to that of a conventional differential measurement system (since $| V_{O1} | + | V_{O2} | \approx cte$ as the target moves), as illustrated in Fig. 4.10.

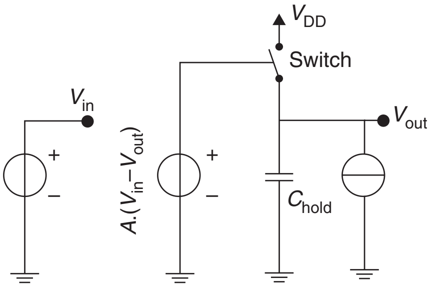
However, in order to exploit the high performance promised by this front-end stage and the ratiometric measurement, a power-efficient, low-offset and linear demodulation has to be performed. A straightforward approach is to utilize a high-speed ADC (e.g., faster than 40 MS/s) and to extract the oscillator output amplitudes in the digital domain. However, such a fast ADC consumes a considerable amount of power, which is not acceptable in a low-power ECS application. Next, we discuss alternative demodulation solutions that can be adapted for the front-end stage and ratiometric measurement.

4.5.1 Precision peak-detection-based ECS interface

Peak detectors employing p-n junction diodes are well-known means by which to achieve AM demodulation.[22] However, due to the temperature dependence of the I-V characteristic of the p-n junctions, such peak detectors



4.10 Transfer characteristic example with ratiometric readout function (using differential sensors with a 6 mm outer radius of the coils).

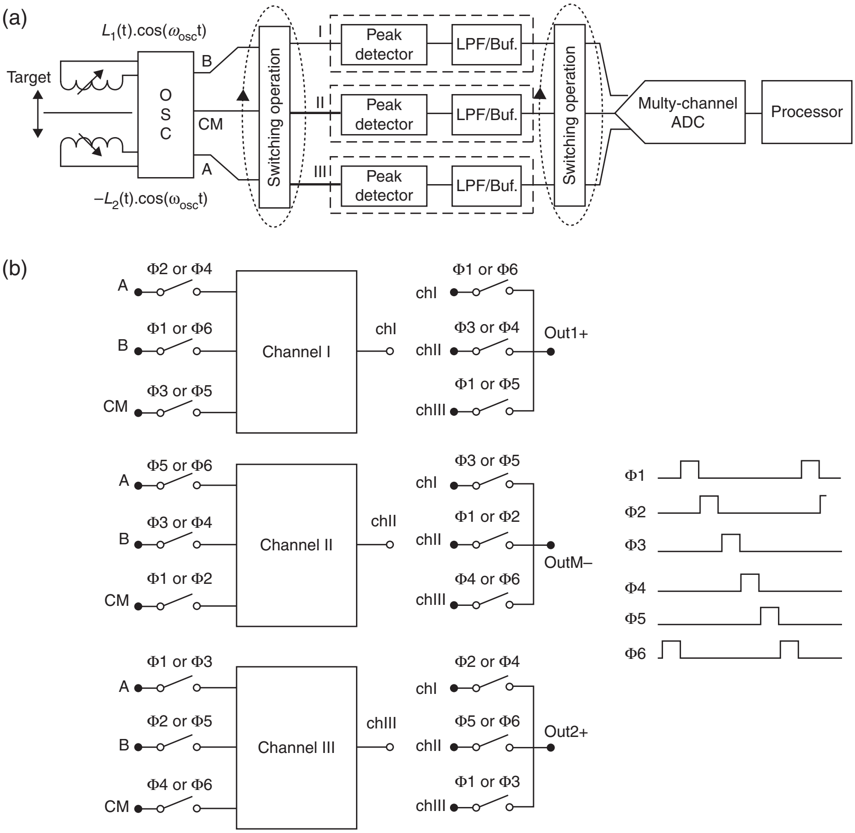


4.11 Precision peak detection principle.

are associated with significant signal dependent thermal drifts.[21,23] To circumvent this problem, an amplifier (or comparator) can be substituted for the p-n junction, as depicted in Fig. 4.11.[24] In this circuit, the amplifier compares the input signal with the capacitor voltage so that, if the input voltage is greater than the capacitor voltage, the amplifier turns on the switch to further charge the capacitor. Consequently, the hold capacitor C_{hold} is charged such that the output voltage becomes equal to the input voltage peaks. In order to follow the input peaks when the input signal amplitude decreases, a small current source in parallel with the capacitor can be considered as a means to discharge it.

Using this principle, Fig. 4.12a shows a demodulator that is adapted for the front-end stage depicted in Fig. 4.9.

The demodulator employs three channels, including precision peak detectors followed by lowpass filters (which also act as buffers). A dedicated switched-channel operation (similar to the dynamic element matching method) is utilized to suppress the effect of associated offsets originating from mismatches in the channels. With this technique, each operation cycle involves six switching phases, in which the oscillator output voltages (A, B and CM in Fig. 4.9) are sequentially interchanged between the inputs of the demodulator channels. At each phase, two channels are connected to the oscillator outputs, while the third detects the common-mode voltage of the oscillator outputs (node CM; see Fig. 4.12b). In this way, the effect of oscillator power supply drift is cancelled out. An ADC which is synchronized with this operation converts the demodulator analog outputs into digital codes for post-processing, including the formation of the ratiometric function. The ratiometric readout function at the end of each phase is defined as:


(a)
L1(t).cos(ωosct)
Target
B
O S C
CM
A
−L2(t).cos(ωosct)
Switching operation
I
II
III
Peak detector
LPF/Buf.
Peak detector
LPF/Buf.
Peak detector
LPF/Buf.
Switching operation
Multy-channel ADC
Processor
(b)
Φ2 or Φ4
A
Φ1 or Φ6
B
Φ3 or Φ5
CM
Channel I
chI
Φ1 or Φ6
chI
Φ3 or Φ4
chII
Φ1 or Φ5
chIII
Out1+
Φ5 or Φ6
A
Φ3 or Φ4
B
Φ1 or Φ2
CM
Channel II
chII
Φ3 or Φ5
chI
Φ1 or Φ2
chII
Φ4 or Φ6
chIII
OutM−
Φ1 or Φ3
A
Φ2 or Φ5
B
Φ4 or Φ6
CM
Channel III
chIII
Φ2 or Φ4
chI
Φ5 or Φ6
chII
Φ1 or Φ3
chIII
Out2+
Φ1
Φ2
Φ3
Φ4
Φ5
Φ6

4.12 (a) Conceptual illustration of the precision peak detection based ECS system adapted for the front-end LC oscillator (Fig. 4.9), which utilizes a switched-channel offset cancellation operation; (b) switching diagram of the offset cancellation operation.

$$D_{\text{out,dmt},\Phi 1} = \frac{(V_A - V_{CM}) - (V_B - V_{CM})}{(V_A - V_{CM}) + (V_B - V_{CM})} = \frac{L_1 - L_2}{L_1 + L_2} \qquad [4.12]$$

where V_A, V_B and V_{CM} are the voltages at the corresponding nodes in Fig. 4.9. If the offsets of the channels are taken into account, one can write:

$$D_{\text{out,dmt},\Phi 1} = \frac{[(V_A + V_{\text{offsetIII}}) - (V_{CM} + V_{\text{offsetII}})] - [(V_B + V_{\text{offsetI}}) - (V_{CM} + V_{\text{offsetII}})]}{[(V_A + V_{\text{offsetIII}}) - (V_{CM} + V_{\text{offsetII}})] + [(V_B + V_{\text{offsetI}}) - (V_{CM} + V_{\text{offsetII}})]} \qquad [4.13]$$

where V_{offsetI}, V_{offsetII} and $V_{\text{offsetIII}}$ are the total input-referred offsets (and low-frequency noise) of the corresponding channels depicted in Fig. 4.12b. With offsets in the range of a few millivolts and oscillator output voltages above a few hundred mV, Expression [4.13] can be modified, using the Taylor series expansion, to:

$$D_{\text{out,dmt},\Phi 1} \cong \frac{V_A - V_B}{V_A + V_B - 2V_{CM}} \cdot \left(1 - \frac{V_{\text{offsetIII}} - V_{\text{offsetI}}}{V_A - V_B}\right)\left(1 - \frac{V_{\text{offsetI}} + V_{\text{offsetIII}} - 2V_{\text{offsetII}}}{V_A + V_B - 2V_{CM}}\right) \qquad [4.14]$$

and at phase Φ2 we have:

$$D_{\text{out,dmt},\Phi 2} \cong \frac{V_A - V_B}{V_A + V_B - 2V_{CM}} \cdot \left(1 - \frac{V_{\text{offsetI}} - V_{\text{offsetIII}}}{V_A - V_B}\right)\left(1 - \frac{V_{\text{offsetI}} + V_{\text{offsetIII}} - 2V_{\text{offsetII}}}{V_A + V_B - 2V_C}\right) \qquad [4.15]$$

Likewise:

$$D_{\text{out,dmt},\Phi 3} \cong \frac{V_A - V_B}{V_A + V_B - 2V_{CM}} \cdot \left(1 - \frac{V_{\text{offsetIII}} - V_{\text{offsetII}}}{V_A - V_B}\right)\left(1 - \frac{V_{\text{offsetII}} + V_{\text{offsetIII}} - 2V_{\text{offsetI}}}{V_A + V_B - 2V_{CM}}\right) \qquad [4.16]$$

$$D_{\text{out,dmt},\Phi 4} \cong \frac{V_A - V_B}{V_A + V_B - 2V_{CM}} \cdot \left(1 - \frac{V_{\text{offsetI}} - V_{\text{offsetII}}}{V_A - V_B}\right)\left(1 - \frac{V_{\text{offsetII}} + V_{\text{offsetI}} - 2V_{\text{offsetIII}}}{V_A + V_B - 2V_C}\right) \qquad [4.17]$$

$$D_{\text{out,dmt},\Phi 5} \cong \frac{V_A - V_B}{V_A + V_B - 2V_{CM}} \cdot \left(1 - \frac{V_{\text{offsetII}} - V_{\text{offsetIII}}}{V_A - V_B}\right)\left(1 - \frac{V_{\text{offsetIII}} + V_{\text{offsetII}} - 2V_{\text{offsetI}}}{V_A + V_B - 2V_C}\right) \qquad [4.18]$$

$$D_{\text{out,dmt},\Phi 6} \cong \frac{V_A - V_B}{V_A + V_B - 2V_C} \cdot \left(1 - \frac{V_{\text{offsetII}} - V_{\text{offsetI}}}{V_A - V_B}\right)\left(1 - \frac{V_{\text{offsetI}} + V_{\text{offsetII}} - 2V_{\text{offsetIII}}}{V_A + V_B - 2V_C}\right) \qquad [4.19]$$

The sum of Expressions [4.14] to [4.19] is free of offset voltages. Mathematically, one can consider the output function of the demodulator as:

$$\overline{D_{\text{out,dmt}}} = \frac{1}{g} \cdot \sum_{j=1}^{g} \left(\frac{1}{6} \sum_{i=1}^{6} D_{\text{out,dmt},\phi i} \right)_j \qquad [4.20]$$

Obviously, the average of the outputs after six (or a multiple of six) consecutive phases will be free of offsets. It will be also free from the low-frequency *1/f* noise, if the rate of offset cancellation operation is higher than the *1/f* corner frequency. Note that the averaging parameter *g* can be increased in order to further suppress the thermal noise at the cost of a larger measurement time.

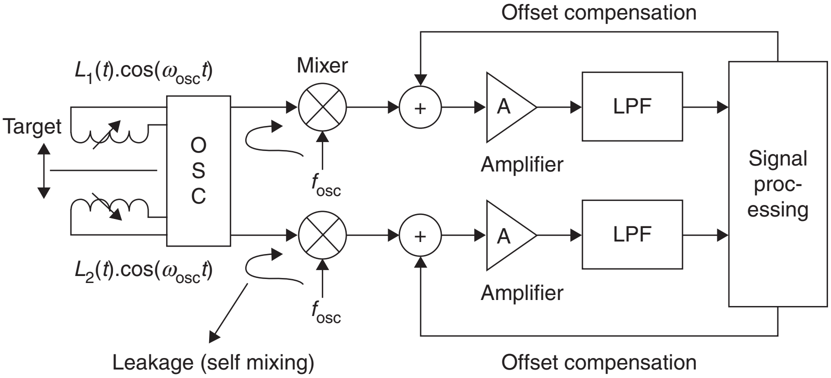
Since the only high-frequency block in the demodulator is an amplifier which drives a very small load (gate of a metal-oxide-semiconductor (MOS) switch), the power consumption in this approach is relatively low. This solution has been realized with a power consumption of less than 10 mW and a signal bandwidth of a few kHz.[4] It is important to mention that in this approach the demodulator requires a relatively wide bandwidth (e.g., ×100 kHz) to perform the offset cancellation operation, which causes the associated noise bandwidth to increase. As a consequence, the obtained resolution is in the medium range (~12 bits). This interface approach is an appropriate candidate for applications with medium resolution, small penetration depth (due to utilizing f_{exc} of 20 MHz), and a relatively low power consumption budget.

4.5.2 Trade-offs in mixer-based interfaces

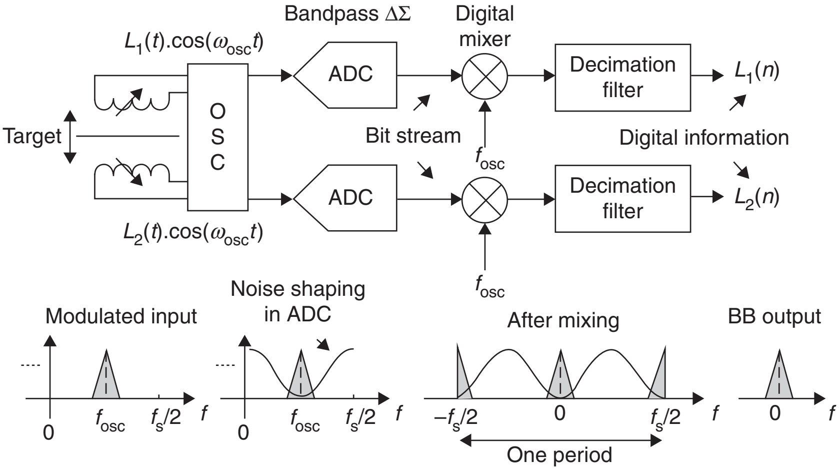
An alternative signal conditioning approach is to utilize a mixer-based demodulator, adapted for the front-end LC oscillator and the ratiometric measurement. In principle, using mixers enables operation in the frequency domain (rather than the time domain, as in peak detectors), which helps to discriminate the frequency bands. This allows the interface to filter out the out-of-band noise (together with external interferences) more efficiently.

In order to make the ratiometric measurement effective on noise suppression, the demodulator (more specifically, the mixer) should be sufficiently linear. Furthermore, in addition to noise considerations, the associated low-frequency errors including drifts and 1/*f* noise have to be suppressed for reasons of stability.

Figures 4.13 and 4.14 show two alternative solutions for the ECS interface. The first solution performs the mixing operation in the analog domain in such a way that the input signal is mixed with a signal frequency that is a


Offset compensation
$L_1(t).\cos(\omega_{osc}t)$
Mixer
Target
O S C
+
A
LPF
Signal proc-essing
Amplifier
f_{osc}
+
A
LPF
$L_2(t).\cos(\omega_{osc}t)$
Amplifier
f_{osc}
Leakage (self mixing)
Offset compensation

4.13 Synchronous detection-based interface with a system level compensation for the offset originated from the self-mixing effect and the mismatches in the amplifier.


Bandpass ΔΣ
Digital mixer
$L_1(t).\cos(\omega_{osc}t)$
ADC
Decimation filter
$L_1(n)$
Target
O S C
Bit stream
Digital information
f_{osc}
ADC
Decimation filter
$L_2(n)$
$L_2(t).\cos(\omega_{osc}t)$
f_{osc}
Modulated input
Noise shaping in ADC
After mixing
BB output
0
f_{osc}
$f_s/2$
f
0
f_{osc}
$f_s/2$
f
$-f_s/2$
0
$f_s/2$
f
0
f
One period

4.14 ECS interface using bandpass ΔΣ converter (a typical choice is $f_s = 4f_{osc}$ where f_s is the ADC sampling frequency).

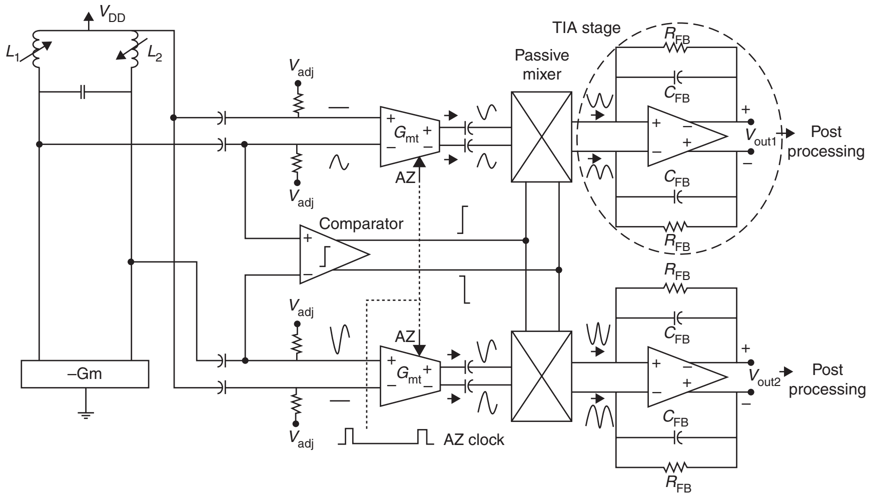
replica of the oscillation frequency (direct conversion). Using the D_{out} readout function, common-mode multiplicative errors that can be introduced by the mixers are eventually suppressed, which makes this solution compatible with the ratiometric measurement method. An imperfection is the offset introduced by the mixers (due to the associated self-mixing effect and charge injection[25]) and the amplifiers (due to mismatches). Provided that an effective offset cancellation operation is performed, this demodulation approach can be the most power-efficient way to realize a high-resolution ECS interface.

Alternatively, a bandpass $\Delta\Sigma$ converter can be employed to digitize directly the narrow-band sensor information that is modulated by the high-frequency carrier signal (Fig. 4.14). The operation principle of the bandpass $\Delta\Sigma$ modulator is based on oversampling and noise-shaping, which is similar to its lowpass counterpart.[26] From a system-level point of view, the main difference is that in a lowpass $\Delta\Sigma$ modulator the noise transfer function (NTF) has zero(s) at DC whereas a bandpass modulator requires zeros at the input frequency so that the noise at this frequency range can be shaped. The output bit-stream of the $\Delta\Sigma$ ADC is applied to a digital mixer to demodulate the AM signal in the digital domain. The signal is then applied to a decimation filter to extract the BB information and to form the ratiometric readout function.

The advantage of this solution is its immunity to DC errors since this type of data converter inherently offers a bandpass characteristic and the mixing operation is performed in the digital domain. However, the power dissipation of this approach is relatively high due to the power needed to provide the noise-shaping characteristic (using active resonators) at a relatively high frequency. More specifically, an overview of state-of-the-art bandpass converters with: ×10 MHz carrier frequency, a dynamic range above 15 bits and a signal bandwidth of a few kHz, shows that the power consumption is in the order of several tens of mW,[27,28] which is too high for an integrated ECS interface in most applications. One solution for this issue is, first, to down-convert the signal to a lower intermediate frequency band (low-IF), and then to apply the signal to a low-IF bandpass $\Delta\Sigma$ modulator. Even though low-IF $\Delta\Sigma$ converters consume a significantly lower amount of power, the interface still requires a preceding linear analog mixer (like in the synchronous detection approach). This implies that if both resolution and power consumption are the leading factors in an ECS interface, then the use of a linear low-noise and power-efficient analog mixer is inevitable.

4.5.3 Synchronous detection-based ECS interface

Based on the demodulation approach shown in Fig. 4.13, a simplified architecture of the ECS interface proposed in Nabavi *et al.* (2010)[29] (including



4.15 Simplified circuit diagram of the synchronous detection-based ECS interface adapted for ratiometric measurement.

two channels) is depicted in Fig. 4.15. Performing the mixing operation in the current domain provides more linearity (than in the voltage domain) since the series on-resistance of the switches cannot introduce distortion to a current signal. Accordingly, the output voltages of the front-end oscillator are converted into fully-differential output currents using two transconductance stages G_{mt}. These stages are capacitively coupled to the oscillator, both to eliminate the DC component of the oscillator outputs and to allow adjustment of the input DC voltage V_{adj} at the desired level. The output currents of the G_{mt} stages are capacitively coupled to passive mixers (or choppers) to demodulate the displacement information. This capacitive coupling prevents any DC output currents from being up-converted by the mixers. This helps to reduce the high-frequency ripples at the output of the demodulator. To synchronize the demodulation with the oscillation frequency, the mixers are driven by a comparator that detects the polarity of the oscillator differential output voltage. Any offset of this comparator appears as a common-mode multiplicative error, which is suppressed by the ratiometric measurement.

Two transimpedance amplifier (TIA) stages convert the demodulated currents into output voltages and pre-filter the high-frequency ripples. The outputs are applied to off-chip RC filters that further limit the bandwidth to a few kHz, which is in line with the signal bandwidth of interest. Finally, the outputs are digitized by an external ADC, after which the ratiometric calculation is performed in the digital domain.

Table 4.2 Qualitative comparison for alternative ECS interface design methods

Specifications interface realization method	LC resonator with SC oscillator[3]	Relaxation oscillator[19]	LC oscillator with ratiometric and precision peak detection[4]	LC oscillator with ratiometric and synchronous detection[29]
Excitation frequency	Low	High	High	High
Dynamic range	Low-medium	Medium-high	Medium	High
Power consumption	Low	High	Low	Low-medium
Bandwidth	High	Medium	Medium	Medium
Stability	Low-medium	Medium	Medium	High

As mentioned, to improve the stability of the interface it is necessary to suppress the effect of low-frequency errors. One source of these errors is the charge injection by the switches in the passive mixers. Since these switches are working at the virtual grounds of the TIAs, the charge injection appears mostly (to the first order analysis) as a common-mode effect that is suppressed at the differential outputs of the TIA stages.

To eliminate the differential component of the charge injection (due to mismatch of the switches) and the self-mixing offset, as well as the offset and the $1/f$ noise of the operational amplifiers (opamps) in the TIAs, an AZ mode is employed. In this mode, by disabling the outputs of the G_{mt} stages, no input signal is applied to the passive mixers. At the same time, the mixers are still driven by the comparator. Therefore, the outputs of the TIA stages correspond to the residual charge injection of the switches and the opamp offsets. This auto-zeroing operation is repeated every 50 ms so that the low-frequency errors can be measured at the outputs and subtracted from subsequent measurements.

This interface achieves a dynamic range of more than 15 bits (better than 100-nm resolution with 3-mm full-displacement scale range) with 1 kHz bandwidth, 30 ppm/°C thermal drift, 18 mW power consumption and an excitation frequency of 20 MHz. It is shown that the utilized AZ operation can improve measurement stability performance by 2–3 bits. This interface is a suitable choice for applications demanding high performance in which the power budget is slightly relaxed.

A standard 0.35 μm Si/Ge BiCMOS technology is utilized for implementing the interface circuits. In such a process, bipolar (NPN) transistors offer a considerably higher transconductance than their nMOS counterparts (as a rule of thumb, a transconductance of about four times larger with a similar biasing current and die area). From a circuit design standpoint (in comparison with complementary MOS (CMOS)), this helps to lower the impedances at the internal nodes significantly, therefore achieving the desired bandwidth with a lower power dissipation. Furthermore, bipolar junction transistors (BJTs) offer a considerably lower 1/*f* noise than their MOS counterparts (about two orders of magnitude in the technology employed, with a similar biasing current and die area). The combination of a low 1/*f* noise and a larger transconductance (under similar conditions) makes the BiCMOS process an appropriate choice for ECS interfaces targeting high excitation frequency, high measurement stability and low-power functionality.

4.6 Conclusion

In this chapter, key design aspects for eddy-current displacement sensors (ECSs) and interfaces were discussed. Furthermore, design challenges related to the design of the sensor and the electronic interface were analyzed and explained. Recent advances in this field were reviewed, compared, and categorized in order to provide the reader with a clear design perspective. The key design points, together with the contents of the chapter, can be summarized as follows:

- The distance to the target sensed by ECSs is defined by three factors: the real sensor-target distance, the eddy currents penetration depth and the target thickness.
- To suppress the imperfections of ECSs (in particular, in precision industrial applications) it is essential to increase the excitation frequency by more than one order of magnitude in comparison to currently popular values of a few MHz. To this end, integrating the sensor and the electronic interface is indispensable.
- The state-of-the-art ECS interfaces are surveyed and their pros and cons are discussed. Table 4.2 provides a qualitative comparison of their performance area.
- Employing second-order LC cross-coupled oscillators is shown to be the most power-efficient way of exciting relatively high-quality factor ECSs.
- Ratiometric measurement can effectively improve the resolution/stability performance of the interface when using a cross-coupled LC oscillator as the front-end stage.
- The synchronous detection approach adapted for an LC front-end oscillator and the ratiometric measurement can be the most power-efficient way to realize a high-performance ECS interface.
- Utilizing BiCMOS technology for implementing an ECS interface provides a proper trade-off in terms of stability, speed and power consumption.

4.7 References

1. Micro-Epsilon, 'Measurement Product Guide 2011', Application note [Online]. 2011. Available at: www.micro-epsilon.com.
2. Lion Precision, 'Eddy-current displacement sensors overview', Application note [Online]. 2011. Available at: www.lionprecision.com.
3. M. Oberle, R. Reutemann, R. Hertle and Qiuting Huang, 'A 10-mW two-channel fully integrated system-on-chip for eddy-current position sensing [in biomedical devices],' *IEEE Journal of Solid-State Circuits*, vol. **37**, no. 7, pp. 916–925 (July 2002).
4. M.R. Nabavi and S. Nihtianov, 'Eddy-current sensor interface for advanced industrial applications,' *IEEE Transactions on Industrial Electronics*, vol. **58**, no. 4, pp. 4414–23 (September 2011).
5. G.Y. Tian, Z.X. Zhao and R.W. Baines, 'The research of inhomogeneity in eddy-current sensors,' *Sensors and Actuators*, vol. **A69**, pp. 148–151 (1998).
6. T. Hitz and S.D. Welsby, 'True position measurement with eddy current technology', *Sensors*, vol. **14**, no. 11, pp. 30–40 (1997).
7. Y. Yating, D. Pingan and W. Zhenwei, 'Study on the electromagnetic properties of eddy current sensor', IEEE International Conference of Mechatronics and Automation, Canada, vol. **4**, pp. 1970–5 (August 2005).
8. S. Fericean and R. Droxler, 'New noncontacting inductive analog proximity and inductive linear displacement sensors for industrial automation', *IEEE Sensors Journal*, vol. **7**, no. 11, pp. 1538–1545 (2007).
9. S.D. Roach, 'Designing and building an eddy current position sensor,' *Sensors*, vol. **15**, no. 9 (1998).
10. A. Weisshaar, L. Hai and A. Luoh, 'Accurate closed-form expressions for the frequency-dependent line parameters of on-chip interconnects on lossy silicon substrate', *IEEE Transactions on Magnetics*, vol. **25**, no. 2, pp. 288–296 (May 2002).
11. K. Kang J. Shi, W. Yin, L. Li, S. Zouhdi, S.C. Rustagi and K. Mouthaan, 'Analysis of frequency- and temperature-dependent substrate eddy currents in on-chip spiral inductors using the complex image method', *IEEE Transactions on Magnetics*, vol. **43**, no. 7, pp. 3243–3253 (July 2007).
12. M.R. Nabavi, Low-power high-performance integrated interface for eddy-current displacement sensors, Delft University of Technology, PhD thesis, ISBN 978-94-6190-750-9, December (2010).
13. I.A. Limanov, 'Analog inductive eddy-current transducers of information,' *Measurement Techniques,* vol. **9**, no. 6, pp. 749–752 (1966).
14. J. Goezinne, 'Electronic circuit for detecting a change relative to a quiescent condition', US Patent No. 6847215 (2005).
15. S. Thoss, O. Machul and B.J. Hosticka, 'A novel architecture for inductive proximity sensors using sigma delta modulation', in Proceedings of the IEEE European Solid-State Circuits Conference (ESSCIRC), Munich, pp. 284–287 (September 2007).
16. X. Chen and T. Ding, 'Flexible eddy current sensor array for proximity sensing', *Sensors and Actuators A*, vol. **135**, no. 1, pp. 126–130 (March 2007).
17. M.R. Nabavi and S. Nihtianov, 'Design strategies for eddy-current displacement sensor systems: review and recommendations', *IEEE Sensors Journal*, vol. **12**, no. 12, pp. 3346–3355 (December 2012).

18. C.C. Enz and G.C. Temes, 'Circuit techniques for reducing the effects of op-amp imperfections: autozeroing, correlated double sampling, and chopper stabilization,' *Proceedings of the IEEE*, vol. **84**, no. 11, pp. 1584–1614 (November 1996).
19. P. Kejik, C. Kluser, R. Bischofberger and R.S. Popovic, 'A low-cost inductive proximity sensor for industrial applications,' *Sensors & Actuators A: Physical*, vol. **110**, pp. 93–97 (2004).
20. D. Vyroubal, 'Eddy-current displacement transducer with extended linear range and automatic tuning', *IEEE Transactions on Instrumentation and Measurement*, vol. **58**, no. 9, pp. 3221–3231 (September 2009).
21. M.R. Nabavi and S. Nihtianov, 'A novel interface for eddy current displacement sensors', *IEEE Transactions on Instrumentation and Measurement*, vol. **58**, no. 5, pp. 1623–1632 (May 2009).
22. A. Bruce Carlson, P. Crilly and J. Rutledge, *Communication Systems*, 4th edn, McGraw Hill (2002).
23. R. G. Meyer, 'Low-Power Monolithic RF Peak Detector Analysis,' *IEEE Journal of Solid-State Circuits*, vol. **30**, no. 1, pp. 65–67 (1995).
24. S.B. Park, J. Wilson and M. Ismail, 'Peak detectors for multistandard wireless receivers', *IEEE Circuits and Devices Magazine*, vol. **2**, no. 6, pp. 6–9 (2006).
25. B. Razavi, *RF Microelectronics*, Upper Saddle River, NJ Prentice Hall (1998).
26. R. Schreier and G.C. Temes, *Understanding Delta-Sigma Data Converters*, Hoboken, NJ Wiley-IEEE Press (2005).
27. P.G.R. Silva, L.J. Breems, K.A.A. Makinwa, R. Roovers and J.H. Huijsing, 'An IF-to-baseband modulator for AM/FM/IBOC radio receivers with a 118 dB dynamic range', *IEEE Journal of Solid-State Circuits*, vol. **42**, no. 5, pp. 1076–1089 (May 2007).
28. T. Yamamoto, M. Kasahara and T. Matsuura, 'A 63 mA 112/94 dB DR IF bandpass ΔΣ modulator with direct feed-forward compensation and double sampling', *IEEE Journal of Solid-State Circuits*, vol. **43**, no. 8, pp. 1783–1794 (August 2008).
29. M.R. Nabavi, M.A.P. Pertijs and S. Nihtianov, 'An interface for eddy current displacement sensors with 15-bit resolution and 20 MHz excitation', in Proceedings of IEEE European Solid State Circuits Conference (ESSCIRC), Seville pp. 290–293 (September 2010).

5

Advanced silicon radiation detectors in the vacuum ultraviolet (VUV) and the extreme ultraviolet (EUV) spectral range

A. GOTTWALD and F. SCHOLZE,
Physikalisch-Technische Bundesanstalt, Germany

DOI: 10.1533/9780857099297.1.102

Abstract: In this chapter, the numerous challenges in reliable measurement of vacuum ultraviolet (VUV) and extreme ultraviolet (EUV) radiation with silicon radiation detectors will be discussed. The most severe issue is the high absorbance of any kind of material in these spectral ranges, which poses stringent requirements on possible detector designs. In particular, a high value of the spectral responsivity and the necessary radiation hardness put contradicting demands on the devices. In order to find a solution, it is necessary to describe and understand in detail the basic physical processes for the interaction between the radiation and the detector. The spectral responsivity of VUV and EUV detectors is measured using monochromatized synchrotron radiation as a radiation source, and electrical substitution radiometers as primary detector standards. We will present selected types of silicon-based semiconductor detectors, and discuss their performance in the context of the named aspects.

Key words: VUV/EUV radiation, silicon photodetectors, spectral responsivity, radiation hardness, VUV/EUV radiation damage.

5.1 Introductory overview

Radiation detection in the VUV and EUV spectral ranges comes with specific physical and technical challenges which are not present in, e.g. the visible and infrared (IR) regimes. However, this part of the electromagnetic spectrum with wavelengths shorter than the air-UV is still of growing interest for industrial applications. The semiconductor photolithographic industry with its process lines at wavelengths of 193 and 13.5 nm requires reliable radiant power measurements and may be seen as a main driving force. Other applications (e.g. UV curing for varnish hardening or water disinfection) also have a certain presence in this field. Moreover, analytical methods used in process control, that is, ellipsometry for the determination of thin film thicknesses are being extended to short wavelengths as well, and also need reliable VUV and EUV detectors.

In the following, we denote the spectral range from 40 to 200 nm (the working range of normal incidence reflective optics) as VUV, and the adjacent shorter wavelength range from 40 nm down to about 1 nm (where grazing incidence mirrors – or, for restricted wavelength intervals, normal incidence multilayer mirrors – are used) as EUV. In both spectral ranges, due to absorption, the radiation must be transported through vacuum. Moreover, particularly in the VUV range, the radiation is almost completely absorbed within a layer of only a few nanometers thickness of any solid material. Due to the existing high photon energies, the radiation is ionizing through the photoelectric effect, and therefore charges insulating materials. This puts stringent requirements on any radiation detecting device. Particularly, the sensitive region of the detector must be as close to the upper surface as possible, and essentially there should be no protecting top layer. Besides detectors based on direct surface photoemission, any detector using the internal photoelectric effect requires a very mature technology to fulfill these requirements. Therefore, silicon-based detectors are the common choice in this field of application. In the following section, we will discuss the basic device solutions to meet the requirements of VUV/EUV detection and present the respective methods for radiometric characterization of the detectors.

5.2 Challenges for radiation detection in the EUV and VUV spectral ranges

In the context of this chapter, we will restrict the discussion to the detection of photons by the photoelectric effect. This means that we will not consider any other excitation processes (e.g. as in photochemical detectors) or purely thermal detection principles. Moreover, we will neither discuss single-photon detectors nor energy dispersive detectors, although most of the considerations presented here will also apply to these detectors, too. The photoelectric absorption creates at least one excited electron in an energy state above the vacuum level. If the absorption occurs close enough to the surface, this electron, as well as secondary electrons generated by electron scattering processes, may escape from the detector to the vacuum, generating an external photocurrent. If the electrons do not escape into vacuum, they populate the conduction band in the detector material. Thus, a first classification of VUV detectors can be based on whether the detected electrons are external or internal.

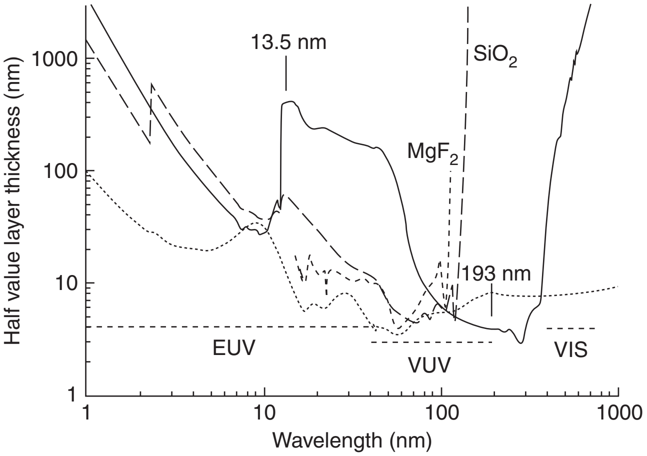
The essential part of external photoemissive detectors is the photocathode, where electrons are ejected by photon excitation. In photoemissive diodes, these electrons are directly measured as a current, whereas in photomultiplier tubes, secondary electrodes (dynodes) are used for a charge multiplication process of the primarily generated electrons. These detectors may be operated in a proportional amplification mode or in a saturated photon counting mode. A well-known example for a photoemissive diode is the so called 'gold diode' where just the external photocurrent from an Au-coated electrode is measured. From the known photoelectric yield of Au in the VUV spectral range,[1] the measured current can be directly related to the radiant power. In metal vacuum chambers, it is principally sufficient to place only the photocathode metal plate insulated from ground into the photon beam and apply a negative bias voltage. The surrounding vacuum chamber, which is usually grounded via other installations like ion pumps and vacuum gauges, acts as the anode. The accuracy of the measurement is limited because the photoelectric yield is already strongly influenced by submonolayers of contamination. However, the method is simple and robust, and the results will at least be correct in the order of magnitude, which is not necessarily assured for detectors based on internal photocurrent measurements if they were subject to prolonged irradiation. Addition of an external anode results in a more sophisticated design and assures collection of the emitted electrons. However, it does not remove the main source of error in the measurement which is the cleanliness of the cathode surface.

On the other hand, detectors using the internal photoeffect offer higher responsivity and potentially higher accuracy, if operated properly. These detectors measure the change in their electrical properties induced by the photogenerated charge. The internal photoeffect excites an electron from an occupied band, inner shell or valence band into a state above the Fermi level. In order to induce any difference in the electronic properties of the detector material, the band into which the electrons are excited must be almost unoccupied such that the electron population induced by the photon irradiation significantly increases the initial population. Ideally, there are no free charge carriers in this band without photon irradiation. Thus, the internal photoeffect leads either to increased conductivity in a homogeneous high-resistance material or to the creation of electron-hole pairs of charge carriers which are then separated by an internal electrical field in a depleted region. Furthermore, the charge carriers created inside the material must be measured at external electrodes. Therefore, not only are materials with a sufficiently long carrier lifetime and an adequate carrier mobility needed, but also some kind of technology for creating contacts to the detector material.

For the photoconductive type of detectors, only one type of material of sufficiently high resistivity of either n- or p-type is needed; no doping technology to create both n- and p-conducting regions to form a p-n junction at the p-to-n interface is required. Therefore, the first examples of detectors from more exotic materials like diamond are, indeed, of the photoconductive type.[2] The use of such wide band-gap materials is mainly triggered by the request to detect very low levels of VUV radiation in the presence of very high levels of visible and IR radiation (e.g. in space-based solar physics applications). These photoconductive detectors, however, do have many drawbacks: an unstable temporal response, limitations in size, homogeneity issues caused by the indispensable contact grids, and a prolonged response time not suited to measure individual pulses of high-frequency pulsed radiation sources.

Any kind of detector used in the VUV and EUV wavelength ranges must be in a vacuum, as the radiation to be detected will be strongly absorbed by any atmosphere. Furthermore, the radiation absorption of all materials in the spectral region around 100 nm is almost equally high, resulting in full absorption of the radiation within the first few nanometres (Fig. 5.1).

No solid material is transparent for wavelengths shorter than 105 nm, which is the absorption edge wavelength of LiF, corresponding to its band-gap energy of 11.8 eV. This particularly means that no window or other transparent layer can be used in front of the detector for protection. For



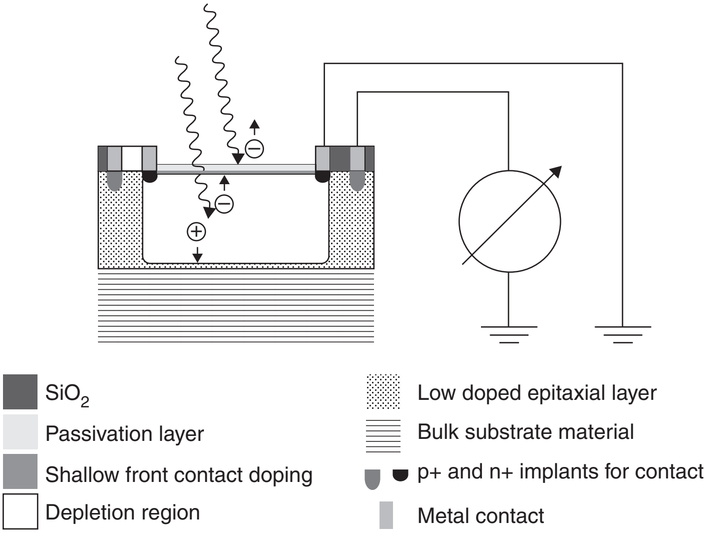
5.1 Half value layer thickness of representative materials in the EUV and VUV wavelength ranges. Examples for UV-transparent, insulating materials are SiO_2 (— — —) and MgF_2 (- - - -). Si (——) is shown as the typical active detector material and Pt (.......) as a representative for a front metal contact layer. The data are taken from Reference 41 for the EUV spectral range and from Reference 42 for the VUV range. The spectral ranges of EUV and VUV as denoted here are also indicated, as well as the most important wavelengths for semiconductor lithography in the respective spectral range. The visible spectral range (VIS) is indicated for reference.

radiation detection with photodiodes, this additionally implies that the sensitive region of the detector must be as close to the upper surface as possible, which severely limits the use of protecting top layers. Another specific issue for the detection of VUV (and even shorter wavelength) radiation is that the photon energy is sufficiently high to create photoelectrons (and respective vacancies) by external photoemission. This is particularly the case in any insulating top layer such as silicon dioxide (SiO_2), which is widely used as the natural passivation layer on silicon. In the UV and visible spectral ranges, SiO_2 is a very appropriate option for a top layer because it is transparent and can, additionally, be used as an antireflection layer. However, vacancies resulting from photo-induced electron emission due to VUV/EUV irradiation will not be quickly re-filled and, as a result, a permanent positive space charge is generated. The resulting electrical field superposes the internal space charge of the p-n junction. For p-on-n diodes, this lowers the junction potential, and may even inverse it at the oxide interface so that detector responsivity drops to zero.

The external photoemission generates a current from the top surface of the diode through the vacuum towards ground. Therefore, two components of the total current must be considered: an external current through vacuum to ground and an internal current towards the reverse of the diode (see Fig. 5.2). For pixel sensors or integrated sensors with several diodes on the same chip, usually the reverse of the diode is used as a common electrode and the contacts on the front are used to read the signal from each individual pixel of the detector. For the VUV spectral range, this approach results in the measurement of the sum of external and internal photocurrent. Any instability or change of the external photocurrent (e.g. due to surface contamination and local charge build-up) will thus be included in the measured signal.

Finally, the most challenging issue for VUV and EUV photodetectors comes from radiation-induced degradation. Three main processes must be considered: (1) charging of the surface layers, (2) the build-up of contamination from the residual gas atmosphere and (3) damage of the surface passivation by prolonged irradiation. The combination of the relatively high photon energies sufficient for the creation of photoelectrons, the extremely short absorption lengths and the need of operation in vacuum results in strong irradiation ageing of detectors which otherwise are regarded as quite stable in other (optical) wavelengths ranges.

To summarize, for the detectors using the internal photoeffect, the external photoemission combined with the strong absorption of radiation in the VUV/EUV spectral ranges causes a significant operational challenge that is not present in applications in the visible and near UV spectral ranges. To meet these challenges, feasible VUV/EUV detectors must have a design adapted to the critical issues presented here.


SiO$_2$
Passivation layer
Shallow front contact doping
Depletion region
Low doped epitaxial layer
Bulk substrate material
p+ and n+ implants for contact
Metal contact

5.2 Scheme of a photodiode with external and internal photocurrent. For p-on-n diodes, both components have the same polarity; for n-on-p diodes the polarity is opposite. The charge separation in the depletion region is shown in the figure for an n-on-p diode. Only if the contact on the front face is put to ground and the current is measured from the bulk side electrode, the internal photocurrent alone is measured.

5.3 Device solutions for radiation detection in the EUV and VUV spectral ranges

The demanding challenges for a stable and highly responsive VUV/EUV detector require devices with a highly uniform depletion zone and shallow cross-section. Therefore, any solution for these challenges in the fabrication of a device must be based on a versatile and mature technology. This strongly favors silicon as the detector material. Silicon technology outperforms any alternative technology – at least, if the integration of further functionality is required to make the sensors smart. For some applications of VUV detectors, however (e.g. solar physics), it is desirable to have a strong suppression of the near UV and visible spectral ranges because the radiant power in these spectral ranges is much higher than that in the VUV. This requirement led to the development of wide band-gap material detectors (known as solar-blind detectors)[2] based on gallium arsenide (GaAs), aluminium nitride (AlN) or even diamond.[2] For these detector materials, however, the current technologies have not reached the same degree of sophistication as in the case for Si. At the moment, this limits the use of those detectors to the detection of very low radiant power (e.g. for astrophysical applications).[3] For the EUV spectral range, metal filters on silicon provide an alternative solution because metals like Zr or Al are somewhat transparent in the wavelength region below 40 nm but, being metals, are opaque to UV and visible radiation.[4,5] As outlined in the previous section, VUV and EUV radiation interacts with the detector by a process of photoelectric absorption close to the surface. For some applications in the field of hard X-ray or IR detection, it may be desirable to hybridize a highly absorbant material like cadmium telluride (CdTe) with silicon for use in the signal read out and amplification.[6] In the VUV and EUV ranges, however, silicon itself is sufficiently absorbant. Therefore, silicon is the first choice for EUV and VUV detectors because it provides low-cost solutions to achieve superior homogeneity, very good electrical properties and long-term stability, as will be discussed in Section 5.5.

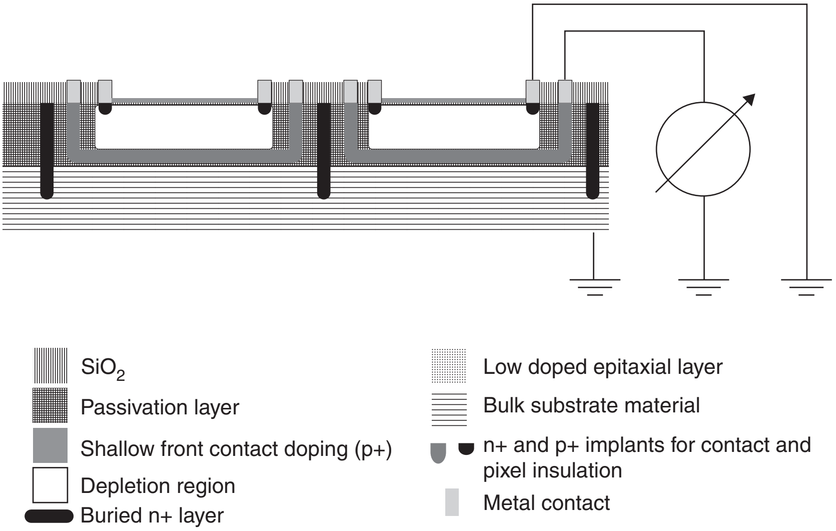
For the visible and near UV spectral ranges, the most common diode structure is a silicon p-on-n junction protected by a SiO_2 passivation layer which is the natural passivation top layer and also acts as an antireflection layer. Because the SiO_2 half value layer thickness (i.e. the thickness at which half of the photons entering the layer are already absorbed by the material) is only 4 nm in the VUV spectral range (Fig. 5.1), an obvious step towards VUV-sensitive diodes was to thin the oxide down to a few nanometers. Initial investigations on the VUV detection properties of silicon diodes were reported in the 1980s.[7] The authors compared inversion layer diodes,[8] n-on-p diodes with P and As doping[9] as well as B-doped p-on-n diodes. Besides issues related to impurity – which were typical issues for diode technology at this time – the surface charging of the oxide was identified as the main issue for device stability. With this consideration in mind, n-on-p doping polarity was suggested because, in this case, the positive surface charge does not counteract the internal field of the junction. Further improvements could be achieved by nitridation of the silicon-oxide interface.[10] Nevertheless, the stability of these devices is not sufficient for higher radiant power applications. For exposures above about 1 J/cm^2, these diodes still degrade.[11] As a result, devices without oxide passivation layer were developed. For the n-on-p type devices, the SiO_2 layer was substituted with a metal-silicide layer,[12] yielding devices with much higher stability.

In the VUV spectral range, it is crucial to achieve a monotonous and very steep doping profile in order to extend the build-in electrical field as close to the surface as possible to collect all generated charge – the photons are absorbed close to the surface due to the low penetration depth. Different doping processes had been used to achieve this, using elements like P, As or B as donors or acceptors.[7] In the case of B-doped p-on-n diodes, the junction as produced by standard doping technologies is not sufficiently shallow. This has been studied in detail for the production of energy dispersive EUV detectors.[13] Energy dispersive detectors measure single photons and convert the charge created by each single photon into a voltage pulse which is proportional to the charge created. These pulses are registered with a digital pulse height analyzer and incrementally stored in a respective channel to create a pulse height spectrum, which represents the measured photon energies. Here, photons in the spectral range from 4 to 10 nm are detected individually. Due to the silicon L-absorption edge, the penetration depth of these photons in silicon is even lower than for the spectral range from 12.5 to 60 nm (see Fig. 5.1). Furthermore, the charge created per photon in the energy range above the Si-L edge is sufficient to detect charge losses by recombination for each individual photon in the measured pulse height spectrum. By measuring the pulse height spectrum, one can differentiate between a situation where every photon is registered but only a part of the initially generated charge is collected, and a situation where the full charge is collected for each photon that is detected but a proportion of the photons is lost by absorption in a dead layer on top of the detector. For the measurement of photocurrent, both situations would yield the same result. Therefore, the energy dispersive detection of the low-penetration-depth but rather high-energy photons with photon energies higher than 100 eV is a good means by which to understand the surface-related recombination losses for photons absorbed in the uppermost region of a silicon detector. The relation between the spectral line shape in energy dispersive photon counting and the charge collection efficiency as a function of depth close to the detector entrance contact is presented elsewhere.[13,14]

In the EUV spectral region, each photon generates a number of electron-hole pairs which is proportional to the photon energy. The proportionality factor is $1/W$, where W is the electron-hole pair creation energy.[15] If there is a loss of charge due to recombination at the surface, the charge measured per pulse is reduced. Thus, the measured line shape for mono energetic photons yields direct information on recombination induced signal losses. By this method, it was shown that standard B-implanted junctions did not achieve sufficient field strength close to the surface to pull the electrons away from the surface sufficiently rapidly. For the energy dispersive detectors, the build-in field strength was improved by an additional P-implantation just beneath the B to increase the local field strength[13] close to the surface.

Another approach is 'pure boron' technology (i.e. a boron doped p-on-n diode with no oxide passivation layer).[16] Chemical vapour deposition (CVD) of a thin amorphous boron layer on top of the diode solved the issue of implanted B doping mentioned earlier because it provides a B reservoir at the surface and thus assures a stable and steep monotonous doping profile, which results in a high build-in electric field extending up to the surface, facilitating an efficient charge carrier collection, even close to the boron–silicon interface. A similar type of radiation-hard and highly sensitive boron doped diode with an SiO_2-free top layer was independently developed elsewhere.[17] Finally, also PtSi/n-Si Schottky diodes proved to be extremely radiation hard in the VUV spectral range.[18] However, their shunt resistance is relatively low, which challenges the measurement of a low photocurrent in the picoampere (pA) range. For all approaches, not only the design determines the detector's properties, but also the stability of the manufacturing process, as the characteristics strongly depend on the purity and uniformity of the layers and interfaces.

So far, only single diode devices have been discussed. For many applications, however, either imaging sensors are required, or several small diodes will be integrated on a single chip for advanced detection systems. This requires the readout of several separate signals from the same chip. The natural solution would be to use a common electrode on the reverse and create the requested detectors with individual electrodes on the front face. In this case, however, the signals must be read from the front. This is a common technique for pixelated detectors in the UV and visible spectral ranges, and is also offered for EUV detectors.[19] The problem is that, for this polarity, the external photocurrent is included in the measured signal. This might be acceptable for EUV detection where the internal current is much higher than the external, and thus a small change in the external current does not significantly change the total signal. In the VUV range, however, both current components are about equal. Therefore, another solution is


SiO$_2$
Passivation layer
Shallow front contact doping (p+)
Depletion region
Buried n+ layer
Low doped epitaxial layer
Bulk substrate material
n+ and p+ implants for contact and pixel insulation
Metal contact

5.3 Operational scheme of pixelated diode detectors with top electrode at ground potential. The individual pixels are insulated with respect to the substrate by an additional p-n junction. Note that the substrate has the same polarity as the entrance contact of the diodes in this case. The example shown is for a p-on-n sensitive junction formed in an n-conductive epi-layer on a p-conductive substrate.

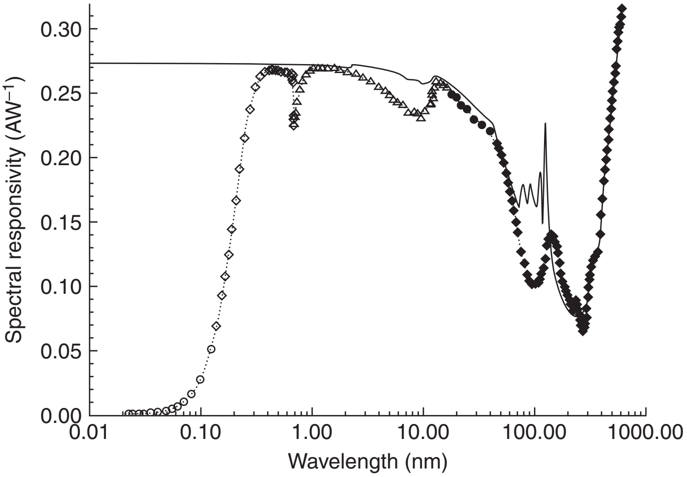
needed. This can be achieved by adding a further p-n junction between the bulk side of the sensitive junction and the substrate of the detector. Then, the sensitive junction can also be individually contacted on its reverse, while the front surfaces of all diodes are connected to a common contact which is put to ground (see Fig. 5.3). There is, however, a price to be paid for this configuration: the additional parasitic p-n junction in parallel to the photodiode will contribute to the dark current and the output capacitance. There are more sophisticated systems based on double-sided wafer processing and charge-drifting systems, such as that for active pixel detectors.[20]

5.4 Methods of radiometric investigation and characterization

In radiometry, the basic physical quantity to be determined is the radiant power of the electromagnetic radiation, which is similar to the radiant flux. Commonly, a radiation detector is described in terms of its spectral responsivity (i.e. its output signal in relation to the incoming radiant power). With the exception of ideal total ('black') absorbers, spectral responsivity is a function of the wavelength. This is particularly the case for silicon-based photodiodes, where the material characteristics (e.g. surface reflectance, internal quantum efficiency) show strong dependencies from the wavelength. In the common event that the radiation has a continuous wave (cw-) characteristic, the current of the silicon photodiode in the short circuit (photovoltaic) mode is usually taken as output. It is also possible to apply bias voltages if needed and, in the event of pulsed radiation, the detector output is then measured in terms of the total charge created, or the voltage pulse height at a certain load resistance – which, for a fixed temporal shape of the pulse, is proportional to the charge generated. The detector's response, however, cannot be assumed to be a unique constant; it may vary across the detectors surface (spatial non-uniformity), with the radiant power (linearity, saturation), temperature and time (degradation by irradiation and long-term temporal instability). Moreover, especially if the measurement is taken under a large angle of incidence, the responsivity is polarization-dependent due to the surface reflectance. For a complete description of a device, all these characteristics must be taken into account.

Basically, any physical measurement must be traceable to the SI (International System of Units) (i.e. the quantity to be measured must be related to the respective realization of the unit: a 'primary standard'). In the case of radiant power, electrical substitution radiometers as primary detectors had already been introduced with the establishment of quantitative radiometry[21] at the end of the nineteenth century. Since the early 1980s, their performance was improved substantially by turning to cryogenic temperatures.[22] Electrical substitution radiometers are thermal detectors based on the equivalence of radiant and electrical heating. By substitution of the absorbed (radiant) power to be measured with electrical power, traceability of the measurement to basic electrical units is established. By definition, the primary standard defines the quantity, and is not to be calibrated. For the long wavelength range (i.e. in the (air-) UV, visible and IR spectral ranges), measurements are commonly realized using (tunable) lasers or monochromatized arc discharges as radiation sources. However, for wavelengths shorter than 200 nm, measurements in air are no longer possible due to the absorption of air-borne oxygen and water. Although it is possible to some extent to use nitrogen (or noble gases) for purging, at least the absorption edge of available window materials gives a definite limit at about 105 nm (see Section 5.2). To cover the whole VUV and EUV spectral range, the use of monochromatized synchrotron radiation for these purposes has been established by certain National Metrology Institutes (NMIs): the Physikalisch-Technische Bundesanstalt (PTB) in Germany,[23] the National Institute of Standards and Technology (NIST) in the United States,[24] and the National Metrology Institute of Japan (NMIJ).[25] Although this is quite an effort from the perspective of instrumentation, the use of synchrotron radiation offers several advantages for radiometric measurements: besides covering a wide spectral range, the radiation characteristics are well-known and the source itself is a high-vacuum machine, providing clean ultra-high vacuum conditions for the measurements. The mutual agreement of the different existing standards of the NMIs mentioned above in their measurement results within their individual uncertainties was proven by an international validation process in recent years.[26–28]

At PTB, synchrotron radiation from the electron storage rings Metrology Light Source (MLS)[29] and BESSY II,[30] both located in Berlin-Adlershof, is used at different beamlines either for the UV/VUV spectral range down to 40 nm[31] or for the EUV spectral range from 40 down to 1 nm, the latter covering the 13.5 nm-wavelength with its relevance for EUV lithographic technology. For these two spectral sub-ranges, two different cryogenic electrical substitution radiometers are also in use at PTB as primary detector standards,[32,33] ensuring the lowest uncertainty possible for the respective range. Both radiometers are equipped with cavity absorbers which ensure almost complete absorption of the radiation in the respective wavelength ranges. Beyond the EUV range, at dedicated X-ray beamlines, it is also possible to extend detector calibration and characterization up to 60 keV photon energy, respectively 0.02 nm (Fig. 5.4).[35] The existing instrumentation allows the detectors to be precisely calibrated with a low level of uncertainties and, furthermore, to be characterized in detail regarding their temporal characteristics, temperature dependence, spatial uniformity, linearity and aging behaviour.



5.4 Scale of spectral responsivity maintained by PTB using monochromatized synchrotron radiation from UV to X-ray regimes as source, and calibrated silicon photodiodes as detectors (traceable to cryogenic radiometers as primary detector standards). The data shown are for an n-on-p silicon diode with a thin SiO_2 passivation layer. The symbols indicate the different beamlines used: Double Crystal monochromator at a wavelength shifter (○), Four Crystal Monochromator at a bending magnet (◇), EUV radiometry beamline (△), Plane Grating Monochromator at an undulator (●), and the Normal Incidence Monochromator at the MLS (◆). The solid line is the model estimation for best-effort silicon radiation detectors. The severe drop in measured responsivity at wavelengths shorter than about 0.2 nm is simply due to the high penetration depth of the X-rays and the only μm-thick sensitive depletion region of the diode.

5.5 Spectral responsivity and radiation hardness of EUV and VUV radiation detectors

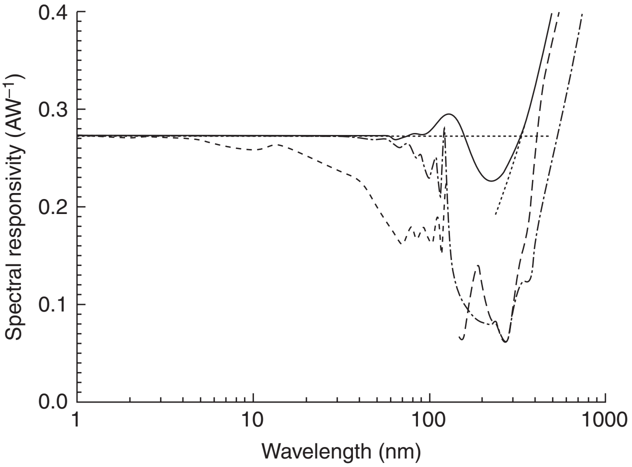
The elementary physical process determining the spectral responsivity of a semiconductor photodetector is the creation of electron-hole pairs in the photodiode depletion zone. The mean energy required to create an electron-hole pair *W* is a material property of any semiconductor. In the UV and visible spectral range (i.e. for wavelengths above 310 nm), it equals the photon energy; there is exactly one electron-hole pair created per absorbed photon – that is, the quantum efficiency ε is constant $\varepsilon = 1$ (electron/photon). At shorter wavelengths (or higher photon energies, respectively) the kinetic energy of the initial photoelectron is sufficiently high to generate secondary electron-hole pairs by electron–electron scattering. For silicon, the probability of these additional ionization processes, however, stays low up to about 4.4 eV photon energy, corresponding to a wavelength of 280 nm.[35] Thus,

$\varepsilon = 1$ holds essentially down to this wavelength and the corresponding value for W increases to $W = 4.4$ eV. With increasing photon energies (i.e. decreasing wavelengths) the secondary ionization process sets in and W converges rapidly to a constant value of W. This asymptotic value is almost reached at 80 nm in the VUV spectral region. For shorter wavelengths (i.e. in the EUV spectral range) the value is constant. For silicon, it has been determined to be $W = (3.66 \pm 0.03)$ eV.[15,35]

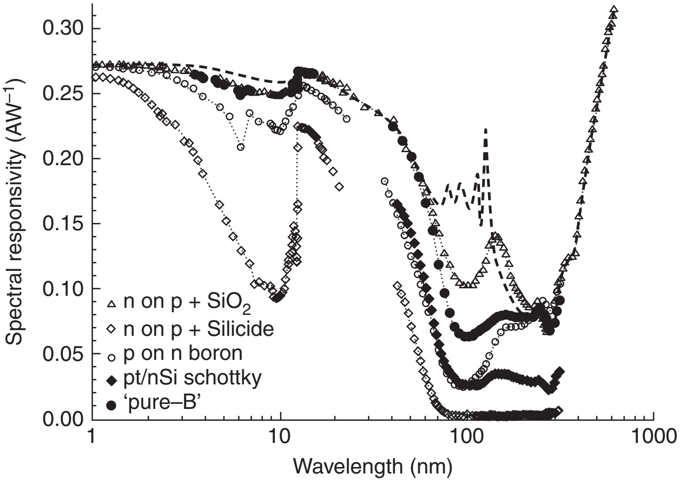
For an ideal detector, the spectral responsivity, s, would only be characterized by this value, $s = e/W$, where e denotes the elementary charge (see Fig. 5.5). To obtain the spectral responsivity of a real detector, additional physical processes must be taken into account. In the EUV, the detector surface does not play any role as the reflectance of all materials goes to zero for wavelengths below about 50 nm. In the VUV spectral range, the reflectance of the detector surface must be taken into consideration. Between 100 and 200 nm, a silicon/silicon oxide surface has a reflectance higher than 50%. Note that, unlike in the UV/visible spectral range (VIS) spectral ranges, the selection of a $\lambda/4$ thickness of the oxide to achieve an antireflection behaviour is not an option due to the strong absorption in the oxide of refractive index $\tilde{n} = n - ik$, with k being much larger than 1. The absorption in (dead) surface layers must be taken into account in all wavelength ranges.

The operation of any semiconductor device requires a passivation of surface electronic states to prevent excessive recombination, at least if the electronically active regions are close to the surface, which is a 'must' for VUV detection. For silicon, this passivation is usually achieved by a thin layer of grown silicon oxide. For improved radiation hardness, however, nitrided oxide or metallic silicide layers are also in use. In the spectral region of highest absorption, around 100 nm, a 5–6 nm silicon oxide layer already absorbs about 80% of the incoming radiation. Further loss mechanisms are due to electron-hole recombination and escape processes, where electrons are lost due to leaving the depletion zone. Although these effects are unavoidable, their individual strength strongly depends on detector quality (e.g. the number of impurity traps or the uniformity of interface layers between depletion and passivation layers). In the VUV spectral range, the absorption length of any kind of material reaches its minimum of, typically, only a few nm, and therefore any change of the detector surface – even of just one monolayer – has a drastic effect. The strong absorption in the surface layers typically reduces the spectral responsivity at 190 nm by about a factor of 5, as compared with the value at 13.5 nm (see Fig. 5.4). Furthermore, the whole energy of the photons is deposited within a very thin layer at the surface of the detector while, at the same time, the detector is highly sensitive to any changes in these uppermost surface layers. This results in strongly enhanced radiation damage in the VUV range in comparison with, for example, the EUV spectral range.

From these considerations, the limiting performance for an ideal silicon photodiode in the EUV/VUV spectral ranges can be estimated, as shown in Fig. 5.5. For short wavelengths below about 10 nm, the intrinsic limit, $s = e/W$, can be almost realized with actual detectors because there are no reflectance losses and a thin passivation layer does not excessively absorb. Above 10 nm, however, the absorptance of even the thinnest passivation layers increases drastically to about 50% in the spectral region around 100 nm. Representative experimental results for the spectral responsivity of different types of silicon photodiodes are shown in Fig. 5.6. Note that the radiation hard devices with metal-silicide passivation have significantly lower responsivity in the VUV spectral range due to the strong absorption of these rather thick passivation layers. They are also less efficient in the EUV spectral range below the silicon L-absorption edge at 12.4 nm. From spectral modelling of the device performance (see e.g. Ref. 15), it can be concluded that



5.5 Responsivity of ideal silicon diodes. The solid line is the limit resulting from the internal quantum efficiency of silicon.[35] The straight dotted horizontal line is the short wavelength approximation with constant $W = 3.66$ eV and the steeply increasing dotted line above 200 nm represents the approximation for the long wavelength range with constant quantum yield $\varepsilon = 1$. The long dash-dotted line includes the loss by external reflection for a 3 nm thick SiO_2 layer, calculated using the data from Palik.[42] This might by tailored in the region where the oxide is transparent by using thicker layers; the behaviour shown is for a 75 nm thick oxide is indicated by the long dashed line. Below 130 nm even a thin oxide begins to absorb. The absorption loss for a 3 nm thick oxide layer is additionally included in the data and is shown by the short-dashed line (below 120 nm, where the oxide becomes transparent, it overlaps with the dash-dotted line accounting for reflection only).



5.6 Absolute spectral responsivity of different types of photodiodes in the VUV and EUV spectral ranges. Shown are data representative of n-on-p diodes with a thin nitrided oxide passivation or a metal-silicide passivation.[19] Boron doped p-on-n[17] and 'pure B'[16] devices achieve responsivity in the range between n-on-p diodes with a thin layer of nitrided oxide and PtSi/n-Si Schottky devices.[18] The theoretical estimation from Fig. 5.4 is also shown as a benchmark for comparison (----). There were no representative data available for the PtSi/n-Si Schottky devices in the EUV spectral range. Note also that the EUV and VUV data are taken for the same type of diode but not always for the same individual device.

this effect is not due to a thin dead layer of silicon but must be explained by a charge collection efficiency monotonously increasing from a starting value below 1 to unity deeper in the active region. This reduced charge collection efficiency for photons absorbed close to the surface hints at charge losses at the silicon–silicide interface. Thus, it indicates the less efficient passivation behaviour of the silicides as compared with SiO$_2$. This effect also contributes to the lower efficiency of silicide-passivated diodes in the VUV spectral range. There, however, it is masked by the strong absorptance of the passivation layer itself.

Regarding radiation hardness, the change of the detector's spectral responsivity under irradiation can be divided in three main categories: growth of surface contamination layers; surface charging of insulating surface layers (like SiO$_2$) due to external photoelectron emission; intrinsic damage (e.g. by creation of trapping states):

1. *Growth of surface contamination layers*: The detectors are used in vacuum because the radiation itself can only propagate in vacuum.

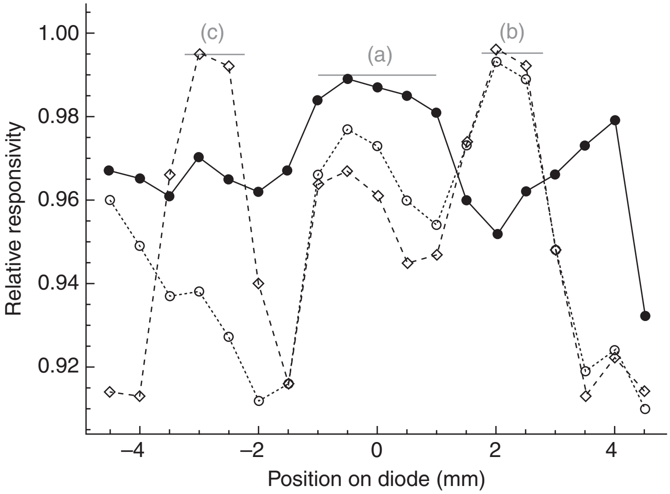
The photon energy, on the other hand, is sufficiently high to ionize the residual gas atmosphere. The most abundant species in a common vacuum system (water and hydrocarbons) are cracked and the resulting radicals either stick to the surface and create carbon contamination or initialize oxidation in case of O-H$^+$. Nitridation of the SiO_2–Si interface turned out to be advantageous for radiation hardness.[10] For prolonged irradiation, however, these oxide passivated diodes degrade extensively.[36,37]

2. *Surface charging of insulating surface layers (like SiO_2) due to external photoelectron emission*: This issue was already addressed in the early investigations of VUV diodes,[7] and changing from p-on-n to n-on-p polarity was suggested as a solution, because in this case the surface charging enhances the intrinsic field of the junction. Particularly, the formation of inhomogeneities was observed when irradiating photo diodes with a small beam (Fig. 5.7a).[11] During an initial time period, the responsivity in the irradiated spot is stable but the region around the spot degrades. This was explained by a negative charging of the adjacent diode region due to photoelectrons accelerated back to the surface by the positive charge inside the irradiation spot and landing nearby on the oxide. This was proven by recovering the initial responsivity with a short irradiation of the initially shadowed area.[11]

   For the VUV spectral range, a pure-boron p-on-n diode was irradiated with a small beam. As discussed, for this polarity the positive charging of some parasitic oxide (probably grown in pinholes of the boron layer) decreases the responsivity. Figure 5.7b shows the recovery of the responsivity over time due to a surface discharge process. This recovery, within a period of days, also indicates that the effect must be caused by rather small islands of oxide in a conductive matrix because the recovery is even slower for homogeneous oxide layers.

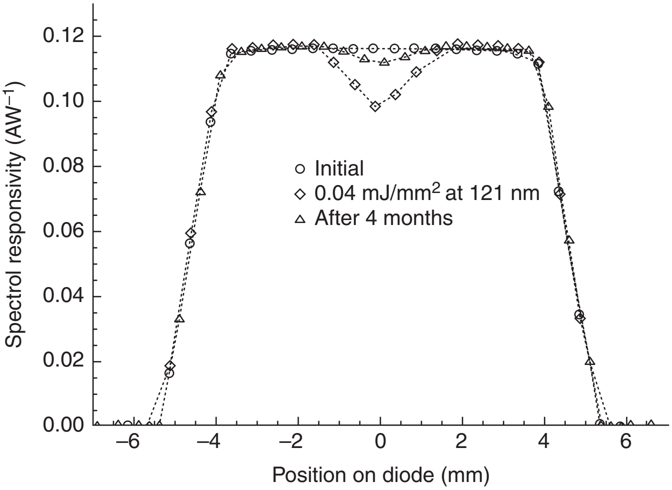
   In practical devices, as shown in Fig. 5.2, the reverse contact forms an outer ring around the sensitive region and the inner contact ring for the face of the electrode. It is connected by deep implanted doping to the reverse of the diode. The insulation between these two electrodes is usually achieved with SiO_2. Irradiation of this region also charges the oxide which, in this case, mainly results in an increased dark current.[38]

3. *Intrinsic damage (e.g. by creation of trapping states)*: This type of damage can be understood as a persistence of the initial impact of the primary process of charging (i.e. the creation of dangling bonds or other types of activated centres in the passivation layer of the diodes). This may result in permanent recombination centres or facilitate oxidation in the presence of O-H radicals (i.e. residual water vapour in an unbaked vacuum system). The damage of the nitrided interface for a



5.7 Effect of local surface charging by narrow photon beams on the responsivity of oxide passivated n-on-p diodes.[11] The diode was irradiated in the centre (position a) with a dose of the order of 10 mJ/mm^2 at a wavelength of 9 nm. The radiation-induced charging of the oxide is positive in the spot (photoelectrons are ejected) and negative around the spot due to electrons back accelerated to the diode by the positive charge and landing nearby. The first measurement (●) was taken three months after this exposure and the responsivity was normalized to the initial value. It was almost stable in the spot but decreased in the adjacent areas. This can be explained by the influence of the positive respectively negative charge on the responsivity of n-on-p diodes; here, a positive charge supports the charge collection because it attracts the electrons towards the n-layer at the surface. A second short exposure of 0.3 mJ/mm^2 was placed at position b and a second scan was taken (○). In the irradiated spot, the responsivity was fully recovered by compensating the negative charge. In the adjacent areas, the responsivity was decreased by negative charging. In the position of the initial spot, the effect was less pronounced because the added negative charge was partly compensated by the remaining positive charge from the first exposure. This was proven by a third exposure at position c and a subsequent scan across the sensitive area (◇).

high irradiation dose could, indeed, be correlated to a radiation-induced oxidation of the surface as the K-absorption edge of O becomes more prominent in the spectral responsivity curve[11]. It is supposed that the combination of radiation-induced dangling bonds at the oxide–silicon interface in combination with O-H+ radicals facilitates surface oxidation. Consequently, the oxide was substituted by an alternative passivation layer. Contamination was particularly investigated in the context of space-based VUV spectroscopy.



5.8 VUV exposure of a ‘pure-B’ diode. Initially, the responsivity was homogeneous across the sensitive region (○). The second scan was taken directly after a short exposure of only 0.04 mJ/mm^2 with 121 nm radiation. The observed decrease in responsivity is almost 20% relative (◇). Here, the positive charge in the spot decreases the charge collection efficiency because the diode is p-on-n polarity. For this type of diode, the initial responsivity was almost recovered after 4 months storage of the device (△).

Whereas the changes due to categories 1 and 2 can, in principle, can reversed by suitable measures (cleaning, discharging), the processes due to category 3, lead to irreversible, permanent damage and, moreover, also influence the electrical properties of the device – typically, significantly raising the dark current,[38] or increasing the surface recombination rate and, hence, leading to reduced responsivity.[39] One common strategy to obtain higher radiation hardness regarding the latter is the use of a conductive (metallic) top layer. However, if this layer is simply added to the diode on top of the usual passivation layer, it substantially reduces the VUV responsivity due to its absorption. Therefore, the oxide passivation itself is replaced by the metallic top layer. At present, the front surfaces of devices with a metal-silicide passivation are significantly thicker than the thinnest oxide layer passivation; also, their passivation properties are inferior. Therefore, these devices have much lower responsivity in the VUV spectral range than diodes with oxide passivation. The latter, however, are not stable under VUV irradiation. Until now, only devices with a highly doped B-layer technique have combined radiation hardness with high responsivity.[17,40]

Summarizing, it can be stated that silicon diodes without an oxide passivation layer provide the best radiation hardness in the VUV/EUV spectral ranges. For the measurement of irradiation stability, it is possible to look directly for the signal change during irradiation. However, all three effects (i)–(iii) will always be observed in combination. To distinguish between them, it is necessary to perform measurements of spectral responsivity before and after irradiation, as well as after cleaning processes. For well-defined irradiation spots, this can also be achieved by scanning across the detector surface. In any event, these measurements need a high level of reproducibility to be comparable.

5.6 Future trends

The present status of VUV and EUV detectors based on the internal photo effect can be summarized as follows:

- The devices are usually silicon devices. The only exception is for applications where high VUV sensitivity and simultaneously no responsivity in the UV and visible is required (solar blindness).
- All detectors with insulating top layers have severe issues with radiation hardness. At the moment, only highly boron doped diodes achieve high VUV responsivity and radiation hardness simultaneously.

Future trends will therefore include work on the improvement of the metal-silicide passivation and the development of more reliable solar blind wide-gap semiconductors, probably by applying the hybridization approach with silicon for those devices also.

For silicon devices, the appearance of ever-more functionalized detectors for different applications is directly on the horizon.

5.7 References

1. Henneken, H., Scholze, F. and Ulm, G. (2000), 'Lack of proportionality of total electron yield and soft x-ray absorption coefficient', *J. Appl. Phys.,* **87**, 257–268.
2. BenMoussa, A., Hochedez, J.-F., Schmutz, W. K., Schühle, U., Nesladek, N., Stockmann, Y., Kroth, U., Richter, M., Theissen, A., Remes, Z., Haenen, K., Mortert, V., Koller, S., Halain, J. P., Petersen, R., Dominique, M. and D'Olieslaeger, M. (2003), 'Solar-blind diamond detectors for LYRA, the solar VUV radiometer on Board Proba II', *Exp. Astron.,* **16**, 141–148.
3. BenMoussa, A., Theissen, A., Scholze, F., Hochedez, J. F., Schühle, U., Schmutz, W., Haenen, K., Stockman, Y., Soltani, A., McMullin, D., Vest, R. E., Kroth, U., Laubis, C., Richter, M., Mortet, V., Gissot, S., Delouille, V., Dominique, M., Koller, S., Halain, J. P., Remes, Z., Petersen, R., D'Olieslaeger, M. and Defise, J. M. (2006), 'Performance of diamond detectors for VUV applications', *Nucl. Instrum. Meth. A*, **568**, 398–405.

4. Canfield, L. R., Vest, R., Woods, T. N and Korde. R. (1994), 'Silicon photodiodes with integrated thin film filters for selective bandpasses in the extreme ultraviolet', *Proc. SPIE,* **2282**, 31–38.
5. Seely, J. F., Korde, R., Hanser, F., Wise, J., Holland, G. E., Weaver, J. and Rife, J. C. (1999), 'Characterization of silicon photodiode detectors with multilayer filter coatings for 17–150 A', *Proc. SPIE*, **3764**, 103–109.
6. Rügheimer, T. K., Gebert, U., Michel, T., Anton, G., Seguinot, J. and Joram, C. (2008), 'Experimental demonstration of a hybrid photon detector concept based on the Timepix detector', *Nucl. Instrum. Meth. A*, **595**, 353–358.
7. Korde, R. and Geist, J. (1987), 'Quantum efficiency stability of silicon photodiodes', *Appl. Opt.,* **26**, 5284–5290.
8. Geist, J., Liang, E. and Schaefer, A. R. (1981), 'Complete collection of minority carriers from the inversion layer in induced junction diodes', *J. Appl. Phys.*, **52**, 4879–4881.
9. Korde, R. and Geist, J. (1987), 'Stable, high quantum efficiency silicon photodiodes by arsenic diffusion', *Solid State Electron*, **30**, 89–92.
10. Korde, R., Cable, J. S. and Canfield, L. R. (1993), 'One gigarad passivating nitrided oxides for 100% internal quantum efficiency silicon photodiodes', *IEEE Trans. Nucl. Sci.,* **40**, 1655–1659.
11. Scholze, F., Rabus, H. and Ulm, G. (1996), 'Spectral responsivity of silicon photodiodes: high-accuracy measurement and improved self-calibration in the soft x-ray spectral range', *Proc. SPIE*, **2808**, 534–543.
12. Korde, R., Prince, C., Cunningham, D., Vest, R. E. and Gullikson, E. (2003), 'Present status of radiometric quality silicon photodiodes', *Metrologia*, **40**, S145–S149.
13. Hartmann, R., Hauff, D., Lechner, P., Richter, R., Strüder, L., Scholze, F. and Ulm, G. (1996), 'Low energy response of silicon pn-junction detectors', *Nucl. Instrum. Meth., A*, **377**, 191–196.
14. Scholze, F. and Procop, M. (2009), 'Modelling the response function of energy dispersive x-ray spectrometers with silicon detectors', *X-ray Spectrom.,* **38**, 312–321.
15. Scholze, F., Rabus, H. and Ulm, G. (1998), 'Mean energy required to produce an electron-hole pair in silicon for photons of energies between 50 and 1500 eV', *J. Appl. Phys.*, **84**, 2926–2939.
16. Nanver, L. K. (2011), 'Silicon photodiodes for low penetration depth beams such as DUV/VUV/EUV light and low-energy electrons', *Advances in Photodiodes*, Gian Franco Dalla Betta (ed.), ISBN: 978-953-307-163-3, Rijeka, Croatia: InTech Europe University Campus, Available from: http://www.intechopen.com/articles/show/title/silicon-photodiodes-for-low-penetration-depth-beams-such-as-duv-vuv-euv-light-and-low-energy-electrons, InTech Europe, 51000 Rijeka, Croatia.
17. Aruev, P. N., Kolokolnikov, Yu. M., Kovalenko, N. V., Legkodymov, A. A., Lyakh, V. V., Nikolenko, A. D., Pindyurin, V. F., Sukhanov, V. L. and Zabrodsky, V. V. (2009), 'Characterization of spatial homogeneity of sensitivity and radiation resistance of semiconductor detectors in the soft X-ray range', *Nucl. Instrum. Meth. A*, **603**, 58–61.
18. Solt, K., Melchior, H., Kroth, U., Kuschnerus, P., Persch, V., Rabus, H., Richter, M. and Ulm, G. (1996) 'PtSi-n-Si Schottky-barrier photodetectors with stable spectral responsivity in the 120–250 nm spectral range', *Appl. Phys. Lett.,* **69**, 3662–3664.

19. http://www.optodiode.com/pro_09.html Brochure download at: http://www.ird-inc.com/brochure/IRD2011.pdf, page 15 (Accessed 30 May 2012).
20. Treis, J., Fischer, P., Halker, O., Harter, M., Herrmann, S., Kohrs, R., Kruger, H., Lechner, P., Lutz, G., Peric, I., Porro, M., Richter, R. H., Struder, L., Trimpl, M. and Wermes, N. (2005), 'DEPMOSFET active pixel sensor prototypes for the XEUS wide field imager', *IEEE Trans. Nucl. Sci.*, **52**, 1083–1091.
21. Kurlbaum, F. (1894), 'Notiz über eine Methode zur quantitativen Bestimmung der stahlenden Wärme', *Ann. Phys.,* **287**, 591.
22. Martin, J. E., Fox, N. P. and Key, P. G. (1985), 'A cryogenic radiometer for absolute radiometric measurements', *Metrologia,* **21**, 147–155.
23. Gottwald, A., Kroth, U., Krumrey, M., Richter, M., Scholze, F. and Ulm, G. (2006), 'The PTB high-accuracy spectral responsivity scale in the VUV and x-ray range', *Metrologia*, **43**, S125–S129.
24. Arp, U., Clark, C. W., Farrell, A. P., Fein, E., Furst, M. L. and Hagley, E. W. (2003), 'Synchrotron ultraviolet radiation facility SURF III', *Rev. Sci. Instrum.,* **73**, 1674–1676.
25. Saito, T. and Onuki, H. (1992), 'Design and performance of a beamline for VUV detector calibration', *J. Spectroscopical Soc. Jpn.*, **43**, 385–393.
26. Scholze, F., Vest, R. and Saito, T. (2010), 'Report on the CCPR pilot comparison: spectral responsivity 10 nm to 20 nm', *Metrologia*, **47**, 02001.
27. Gottwald, A., Richter, M., Shaw, P.-S., Li, Z. and Arp, U. (2011), 'Bilateral NIST–PTB comparison of spectral responsivity in the VUV', *Metrologia*, **48**, 02001.
28. Tanaka, T., Kato, M., Kurosawa, T., Morishita, Y., Saito, N., Suzuki, I. H., Krumrey, M. and Scholze, F. (2012), 'First comparison of spectral responsivity in the soft x-ray region', *Metrologia,* **49**, 501–506.
29. Klein, R., Thornagel, R., Ulm, G., Feikes, J. and Wüstefeld, G. (2011), 'Status of the metrology light source', *J. Electron Spectr. Relat. Phenomena*, **184**, 331–334.
30. Klein, R., Krumrey, M., Richter, M., Scholze, F., Thornagel, R. and Ulm, G. (2002), 'Radiometry with synchrotron radiation at the PTB laboratory at BESSY II', *Synchrotron Radiation News,* **15**, 23–29.
31. Gottwald, A., Kroth, U., Richter, M., Schöppe, H. and Ulm, G. (2010), 'Ultraviolet and vacuum-ultraviolet detector-based radiometry at the Metrology Light Source', *Meas. Sci. Technol.,* **21**, 125101, doi:10.1088/0957-0233/21/12/125101.
32. Rabus, H., Persch, V. and Ulm, G. (1997), 'Synchrotron-radiation operated cryogenic electrical-substitution radiometer as high-accuracy primary detector standard in the ultraviolet, vacuum ultraviolet and soft x-ray spectral ranges', *Appl. Opt.,* **36**, 5421–5440.
33. Rabus, H., Klein, R., Scholze, F., Thornagel, R. and Ulm, G. (2002), 'Validation of the uncertainty budget for soft X-ray radiant power measurement using a cryogenic radiometer', *Metrologia*, **39**, 381–389.
34. Gerlach, M., Krumrey, M., Cibik, L., Müller, P. and Ulm, G. (2007), 'A cryogenic electrical substitution radiometer for hard X-rays', *Nucl. Instr. and Meth. A*, **580**, 218–221.
35. Scholze, F., Henneken, H., Kuschnerus, P., Rabus, H., Richter, M. and Ulm, G. (2000), 'Determination of the electron-hole pair creation energy for semiconductors from the spectral responsivity of photodiodes', *Nucl. Instrum. Meth A*, **439**, 208–215.

36. Scholze, F., Brandt, G., Müller, P., Meyer, B., Scholz, F., Tümmler, J., Vogel, K. and Ulm, G. (2002), 'High-accuracy detector calibration for EUV metrology at PTB', *Proc. SPIE*, **4688**, 680–689.
37. Scholze, F., Klein, R. and Bock, T. (2003), 'Irradiation stability of silicon photodiodes for extreme-ultraviolet radiation', *Appl. Opt.*, **42**, 5621–5626.
38. Shi, L., Nanver, L. K. and Nihtianov, S. N. (2011), 'Stability characterization of high-sensitivity silicon-based EUV photodiodes in a detrimental industrial environment', *IECON 2011 – 37th Annual Conference on IEEE Industrial Electronics Society,* 7–10 November 2011, Melbourne, Australia, 2651–2656.
39. Gullikson, E. M., Korde, R., Canfield, L. R. and Vest, R. E. (1996), 'Stable silicon photodiodes for absolute intensity measurements in the VUV and soft x-ray regions', *J. Electr. Spectrosc. Rel. Phenom.*, **80**, 313–316.
40. Shi, L., Sarubbia, F., Nanver, L. K., Kroth, U., Gottwald, A. and Nihtianov, S. (2010), 'Optical performance of b-layer ultra-shallow-junction silicon photodiodes in the VUV spectral range, *Procedia Eng.,* **5**, 633–636.
41. Henke, B. L., Gullikson, E. M. and Davis, J. C. X-ray interactions: photoabsorption, scattering, transmission, and reflection at E=50-30000 eV, Z=1-92, Atomic Data and Nuclear Data Tables Vol. 54 (no. 2), 181–342 (July 1993), online: http://www-cxro.lbl.gov.
42. Palik, E. D. '*Handbook of Optical Constants of Solids*', Academic Press, Inc., 1985, San Diego, New York, Boston, London, Sydney, Tokyo, Toronto.

6

Integrated polarization analyzing CMOS image sensors for detection and signal processing

---

M. SARKAR, Indian Institute of Technology, India and
A. J. P. THEUWISSEN, Delft University of Technology,
The Netherlands

DOI: 10.1533/9780857099297.1.124

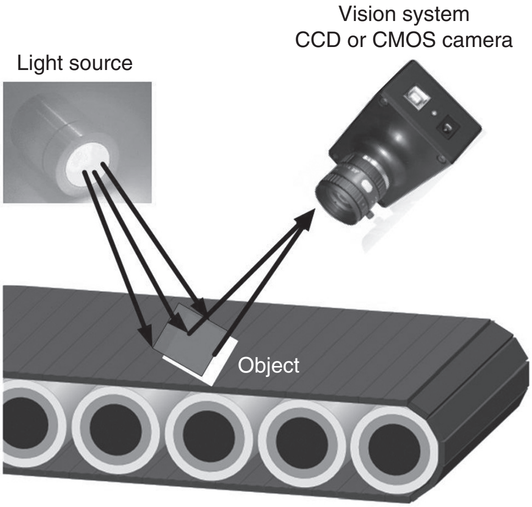
Abstract: This chapter discusses polarization vision for machine vision applications and shows it to be useful in the real time classification of materials into metallic and dielectric surfaces. Light parameters such as intensity, color and polarization are important in industrial machine vision applications to enhance the contrast of the object being imaged. Intensity imaging is not ideal when imaging specular surfaces as they reflect the entire incident light and this often saturates the vision system, while color imaging is influenced by the spectral nature of the surface. Polarization imaging can provide a more general description of light than either the intensity or the color alone, and can therefore provide richer sets of descriptive physical constraints for the interpretation of the imaged scene. Polarization filters help in reducing the transmitted intensity and in imaging transparent objects, and enhance contrast between similar regions. Polarization vision can simplify some visual tasks that are more complicated or practically infeasible when limited to intensity and color processing.

Key words: machine vision, image sensor, complementary metal–oxide–semiconductor (CMOS), polarization vision, metallic wire-grid polarizer, material classification.

6.1 Introduction

Machine vision has become a key technology for detecting and processing signals in industrial applications, and is used to ensure the quality demands of manufactures and customers are met. It requires a light source to illuminate the object of interest and a camera system mounted to image the object. The incident light from the illuminating source is reflected by the object and is captured by the camera system (as shown in Fig. 6.1).

From Fig. 6.1, it is evident that a machine vision system depends not only on the vision system specifications, but also the nature of the light source, the



6.1 Lighting arrangement in machine vision applications.

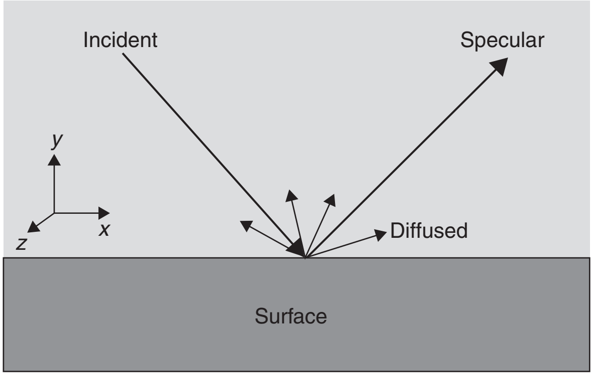
light itself, and the properties of the object. To image a feature of the object of interest, light must come from the light source, reflect off the object, and be collected by the lens. Thus, the properties of light source and the nature of the object influence quality and consistency measurements.

Ideally, machine vision lighting should maximize the contrast on the features on object of interest while minimizing the contrast elsewhere.[1] The object contrast is dependent on the brightness and the spectral content of the light source. The appropriate level of illumination brightness is important: a very bright light source would saturate the vision system while, for features that are difficult to pick out, the level of reflected light needs to be sufficiently great to be detected by the charge-coupled device (CCD) or the complementary metal–oxide–semiconductor (CMOS) imaging system. In a good lighting system, all the light from the source is incident on the object and all the reflected light from the object reaches the imaging system. However, inconsistency due to the presence of ambient light adversely affects the measurements. Together with the brightness of the light source, the direction of the incident light on the object is also important in highlighting features of the object. The spectral content of the light source also plays an important role as objects reflect and absorb colors differently due to their structural arrangements. The contrast of the object being imaged thus depends also on the color of the incident light and the direction in which the light is incident.

There has been no truly vision-specific lighting available in market. The most commonly used lighting sources for machine vision applications are the fluorescent and incandescent types. These ambient light sources are not the ideal choice. A controlled light source that accentuates object features of interest and minimizes distracting object features is to be preferred.[1] The illumination can be focused or diffused. Focused sources provide high light intensity and light efficiency, creating sharp image edges, casting shadows and accenting surface features. However, they create bright reflections and could cause the vision system to saturate. Diffused light covers a large solid angle around the object, thus distributing the intensity and preventing saturation of the vision system, but causes blurred edges in images.

The contrast detected by the imaging system also depends on the reflection/absorption properties of the object surface. The object surface shape, geometry, composition, color and absorption/reflection behavior will determine the contrast perceived by the imaging system. A specular surface will reflect a directional light source differently from a diffused surface. Specular reflection is a mirror-like reflection from a surface in which light from a single incoming direction is reflected into a single outgoing direction (as shown in Fig. 6.2). The specular reflection is governed by the laws of reflection wherein the reflected angle is equal to the incident angle. In diffused reflection, the incident light is scattered over a range of outgoing angles (as shown in Fig. 6.2). A diffused reflection is caused by the reflected rays from internal scattering inside the surface of the medium. When light strikes a surface interface, a part of the light passes through the boundary. This light is re-emitted randomly after undergoing internal scattering.

Specular reflection is bright, as almost the entire light intensity is reflected back. The high brightness of a specular reflection can saturate the vision system. Furthermore, specular reflection is also very sensitive to the direction of the incoming light and is thus is not the preferred option. In contrast, light from a diffused reflection is dim and is not highly-dependent on the



6.2 Reflection types: specular and diffused reflections.

direction of the incident light, thus it is more stable and suitable for machine vision systems. To image specular surfaces which reflect a large portion of the incident light, a diffused light source is used; to image a diffused surface, a focused source is used to reflect back most of the light.

One problem with a specular reflection or with light reflected from a specular surface is that it causes glare. Glare degrades the image contrast and thus is undesirable. Polarization illumination is commonly used to reduce the image glare of specular surfaces.[1] In polarization illumination, polarizing filters are used in pairs, one polarizer is placed in front of the light source, and another (analyzer) is placed in front of the imaging system with its axis perpendicular to the first polarizer. The specularly reflected light retains its polarization and thus is blocked by the analyzer. However, the diffused light is randomly polarized and thus passes through the analyzer, preventing saturation of the imaging system. The polarization of the reflected light also depends on the direction of the incident light and the nature of the reflecting surface. This polarized component can be used to detect the nature of the surface – for example, to discriminate between a metal and a dielectric surface.[2]

State-of-the-art polarization image sensors utilize optical imaging systems that are external to the detectors.[3] These polarization sensors have either single or multi-axis filter arrays which measure the polarization information in real time. Compactness of design and speed of generating polarized images can be greatly enhanced by incorporating an array of microscopic polarization filtering optics directly onto a photosensitive chip. This simplifies the image capturing and makes the system faster. Thus, polarization illumination together with image sensors with embedded linear polarizers mean that the machine vision system is autonomous and faster. They are therefore better-suited for imaging objects in machine vision applications and also for real time material classification.

6.2 Polarization vision

The three basic characteristics of light are intensity, wavelength or color and polarization. The intensity or brightness is an expression of the amount of light emitted from the surface of the light source per unit area. The colors of the visible spectrum stretch from violet, with the shortest wavelength, to red, with the longest wavelength. Polarization is the plane of vibration of the electric field vector of light.

According to Bernard and Wehner,[4] 'polarization vision' is the ability to discriminate between two monochromatic lights of the same intensity, but with a different angle of polarization and/or degree of linear polarization (DOLP). Polarization vision is a generalization of intensity vision and analogous to color vision. Color vision is useful in segregating scenes and in recognition of objects, while polarization vision discriminates reflectances from a scene of similar brightness using the difference in their spectral features.[4] The study of polarization is more general than that of intensity or color and can simplify some of the visual tasks (e.g. region and edge segmentation, material classification, etc.) which are more complicated or practically infeasible when limited to intensity and color information. As intensity is the compression of polarization information, a polarization camera can also function as a conventional intensity camera.

Virtually all light that is reflected in most environments is partially polarized, which means that it can be represented by the sum of unpolarized states and completely linear polarized states. Even though the sun itself produces fully unpolarized light, partially polarized light is abundant in natural scenes.[5] In this chapter, the term 'polarized light' will refer to partially polarized light.

Polarization-dependent optical systems generally fall into three types:

1. systems which measure polarization properties and are therefore sensitive to any system-induced polarization variation;
2. systems which accurately measure intensities and are therefore sensitive to transmissions which change with polarization; and
3. systems which route light based on its polarization and are therefore sensitive to any system-induced variation in polarization.[6]

A polarization-dependent optical system similar to type 2, which measures the intensities varying with polarization, has been designed and is presented[7] and briefly described in Section 6.4.

6.2.1 Polarization of light: basics

Light, viewed classically, can be considered as a transverse wave wherein the underlying oscillations of the electrical and magnetic fields occur along directions perpendicular to the direction of propagation of the wave. Longitudinal waves, in contrast, have oscillations confined in the direction of the propagation, as in sound waves. Polarization is a phenomenon specific to transverse waves. It is the distribution of the electrical field in the plane normal to the propagation direction. An unpolarized or randomly polarized electromagnetic wave is one in which the orientation of the electric vector changes randomly.

The mathematical representation of a plane wave propagating in the z direction is:

$$\vec{E} = E_0 \cos(kz - \omega t + \varphi_0) \qquad [6.1]$$

where E_0 is the amplitude, k is the propagation (or wave) constant ($k = 2\pi/\lambda$), ω is the circular frequency ($\omega = kc = 2\pi c/\lambda$), t is time and φ_0 is the initial phase.

The electric vector $\vec{E}$ may be decomposed into parallel E_x and perpendicular E_y components as:

$$\vec{E} = E_x \vec{x} + E_y \vec{y} \tag{6.2}$$

where E_x and E_y are expressed in the form:

$$E_x = E_{x0} \cos(kz - \omega t + \varphi_{x0}) \tag{6.3}$$

$$E_y = E_{y0} \cos(kz - \omega t + \varphi_{y0}) \tag{6.4}$$

From Equations [6.3] and [6.4] we have:

$$\frac{E_x}{E_{x0}} = \cos(\zeta)\cos(\varphi_{x0}) - \sin(\zeta)\sin(\varphi_{x0}) \tag{6.5}$$

$$\frac{E_y}{E_{y0}} = \cos(\zeta)\cos(\varphi_{y0}) - \sin(\zeta)\sin(\varphi_{y0}) \tag{6.6}$$

where $\zeta = kz - \omega t$.

From Equations [6.5] and [6.6] we obtain:

$$\left(\frac{E_x}{E_{x0}}\right)^2 + \left(\frac{E_y}{E_{y0}}\right)^2 - 2\left(\frac{E_x}{E_{x0}}\right)\left(\frac{E_y}{E_{y0}}\right)\cos(\delta) = \sin^2(\delta) \tag{6.7}$$

where $\delta = \varphi_{x0} - \varphi_{y0}$ is called the *phase shift*.

Equation [6.7] defines an ellipse, thus the polarized electromagnetic wave of this form is known as an *elliptically polarized* wave. The polarization ellipse is very useful since it allows the description of all states of a completely polarized light wave using a single equation. However, this representation is inadequate because the light vector propagating in space traces an ellipse or any special form of an ellipse (e.g. the circle or a straight line) in a time interval of the order of 10^{-15} s, which is very short to be able to trace the ellipse. Furthermore, the polarization ellipse only represents a completely polarized light and cannot be used to describe either unpolarized light or partially polarized light.

The circular form of the polarization ellipse can be obtained if the phase shift satisfies $\delta = n\pi/2$, where ($n = \pm 1, \pm 3, \ldots$) and $E_{x0} = E_{y0} = E_0$ then $\sin(\delta) = \pm 1$ and $\cos(\delta) = 0$, thus Equation [6.7] becomes:

$$E_x{}^2 + E_y{}^2 = E_0{}^2 \qquad [6.8]$$

Equation [6.8] defines a circle, thus the polarized electromagnetic wave of this form is known as a *circular polarized* wave.

The linear form of the polarization ellipse is obtained if the phase shift $\delta = n\pi$, where ($n = 0, \pm 1, \pm 2, \ldots$), then $\sin(\delta) = 0$ and $\cos(\delta) = \pm 1$, thus Equation [6.7] becomes:

$$\left(\frac{E_x}{E_{x0}} + \frac{E_y}{E_{y0}}\right)^2 = 0 \; or \; E_y = \pm \frac{E_{y0}}{E_{x0}} E_x \qquad [6.9]$$

Equation [6.9] defines a straight line, thus the polarized electromagnetic wave of this form is known as a *linear polarized* wave. The electrical field of a linearly polarized wave vibrates in the same direction at all times at any particular point. A linearly polarized light is often referred to as *plane polarized light* or simply *polarized light*.

Linear polarization can be of two types, vertically or *s*-polarized* and horizontally or *p*-polarized. A vertically polarized wave is one for which the electrical field lies only in the *x–z* plane (as shown in Fig. 6.3a). A horizontally polarized wave is one where the electrical field lies only in the *y–z* plane (as shown in Fig. 6.3b).

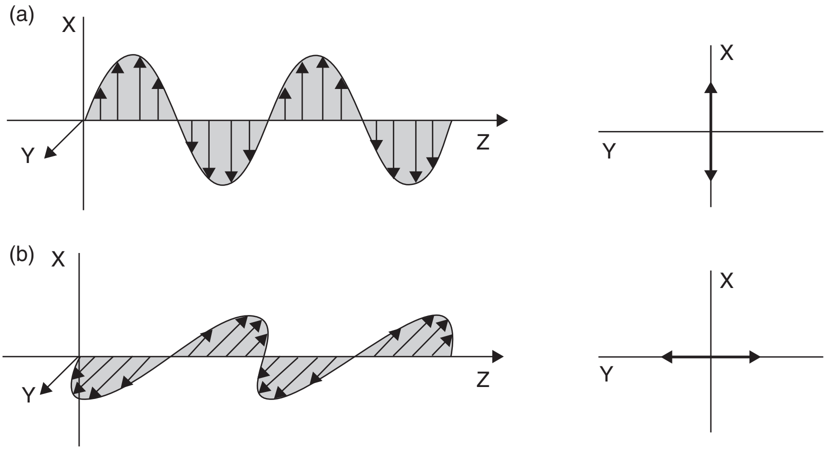
Circular polarized light is rare in nature[8] and thus the discussion in this chapter will be limited to the detection and analysis of linear polarization.

6.2.2 Polarization of light from an unpolarized beam

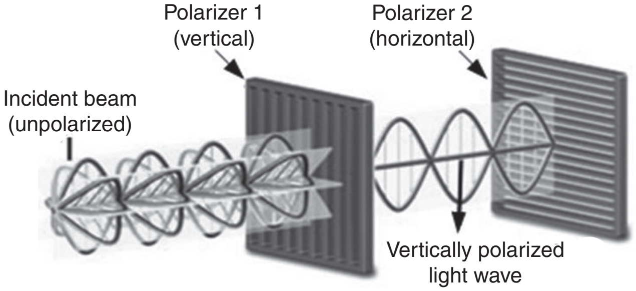
A linearly polarized beam is obtained from an unpolarized beam by removing all waves from the beam except those the electrical field vectors of which oscillate in a single plane. A polarized light from an unpolarized light can be achieved through one of the following phenomena: absorption, reflection, refraction or scattering.

---

* The name *s*-polarized comes from the German word '*senkrecht*' (meaning 'perpendicular') and *p*-polarized is parallel polarized with respect to the plane of incidence.


(a)
X
Y
Z
X
Y
(b)
X
Y
Z
X
Y

6.3 Linear polarization: (a) *s*-polarized wave and (b) *p*-polarized wave.


Polarizer 1
(vertical)
Polarizer 2
(horizontal)
Incident beam
(unpolarized)
Vertically polarized
light wave

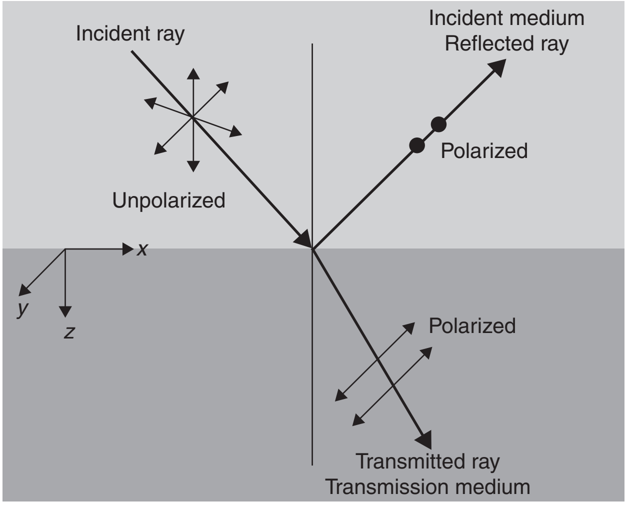
6.4 Polarization by absorption.[9]

Polarization by absorption

The most common technique for producing polarized light is by passing the unpolarized light through a material that absorbs electrical field vibrating in any direction other than the vibrations in a plane parallel to a specified direction. Such a device is known as a *polarizer*. One such type of polarizing filter was invented by Edwin H. Land in 1928 and was called a Polaroid. A Polaroid consists of long-chain organic molecules aligned in one direction, placed in a plastic sheet. They form a closely spaced linear grid, which restricts the passage of light to only that vibrating in the same direction as the grid while absorbing the vibrations in all other directions (as shown in Fig. 6.4).

Polarization by reflection

When an unpolarized light beam is reflected from a surface, the reflected light may be completely polarized, partially polarized, or unpolarized, depending

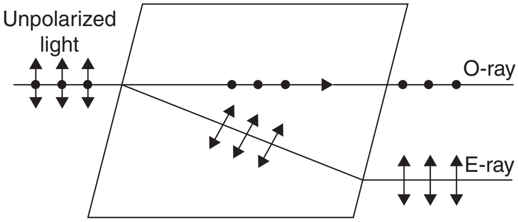


6.5 Polarization by reflection.

on the angle of incidence. For an angle of incidence of 0°, the reflected light is unpolarized, while for other values of angle of incidence, the reflected light is polarized with a vibration direction parallel to the reflecting surface (as shown in Fig. 6.5). If part of the incident light enters the material, the refracted light will also be polarized with a vibration direction perpendicular to the path of the refracted ray.

Polarization by refraction

Optically anisotropic materials have different optical properties in different directions. Double refraction is an optical property where a single ray of unpolarized light entering an anisotropic medium is split into two rays, each traveling in a different direction (as shown in Fig. 6.6). Such materials with different indices of refraction associated with different crystallographic directions are known as *birefringent* materials. Birefringent uni-axial crystalline materials have a unique axis of symmetry, known as the *optic axis*, where no double refraction occurs. The propagation of light along the optic axis would be independent of its polarization; its electrical field is everywhere perpendicular to the optic axis and it is called the ordinary- or O-ray. A light wave with an electrical field parallel to the optic axis is called the extraordinary- or E-ray. When an unpolarized light enters a birefringent crystal, the ordinary wave will pass straight through, while the extraordinary wave is displaced by a distance depending on the thickness of the crystal and the angle of refraction.



6.6 Polarization by double refraction.

Polarization by scattering

Polarization or depolarization also occurs when light is scattered while traveling through a medium. When light strikes the atoms of a material, it often excites the electrons of those atoms, causing them to vibrate. The vibrating electrons then produce their own electromagnetic wave which is radiated outwards in all directions. This newly generated wave strikes neighboring atoms, forcing their electrons to vibrate at the same original frequency. These vibrating electrons produce a further electromagnetic wave which, again, radiates outwards in all directions. This absorption and re-emission of light waves causes the light to be scattered about the medium. This scattered light is partially polarized.

6.2.3 Polarization of light: representation

The polarization ellipse discussed in Section 6.2.1 describes the polarized light in terms of its amplitude. The amplitude of the optical field cannot be observed, but it is possible to observe and measure its intensity, which is equal to the time average of the square of the field amplitudes. This can be done by taking a time average of the time dependent quantities in Equation [6.7]. The time average of the field components are defined by the following equation:

$$\left\langle E_x(z,t)E_y(z,t)\right\rangle = \lim_{T\to\infty}\left(\left(\frac{1}{T}\right)\int_0^T E_x(z,t)E_y(z,t)dt\right) \qquad [6.10]$$

where T is the time of the measurement. To evaluate Equation [6.7], the denominator is first removed by multiplying by the factor $4E_{0x}^2 E_{0y}^2$. We then have:

$$\left(4E_{y0}^2\right)\left\langle E_x^2(z,t)\right\rangle+4E_{x0}^2\left\langle E_y^2(z,t)\right\rangle-8E_{x0}E_{y0}\left\langle E_x(z,t)E_{yx}(z,t)\right\rangle\cos\delta=\left(2E_{x0}E_{y0}\sin\delta\right)^2 \qquad [6.11]$$

Using Equations [6.3], [6.4] and [6.10], the time averages in Equation [6.11]are found to be:

$$\left\langle E_x(z,t)\right\rangle = \tfrac{1}{2}E_{x0}{}^2$$

$$\left\langle E_y(z,t)\right\rangle = \tfrac{1}{2}E_{y0}{}^2 \qquad [6.12]$$

$$\left\langle E_x(z,t)E_y(z,t)\right\rangle = \tfrac{1}{2}E_{x0}E_{y0}\cos\delta$$

Substituting Equation [6.12] in [6.11] and rearranging we obtain:

$$\left(E_{x0}^2 + E_{y0}^2\right) - \left(E_{x0}^2 - E_{y0}^2\right) - \left(2E_{x0}E_{y0}\cos\delta\right)^2 = \left(2E_{x0}E_{y0}\sin\delta\right)^2 \qquad [6.13]$$

From Equation [6.13] we can define:

$$S_0{}^2 = S_1{}^2 + S_2{}^2 + S_3{}^2 \qquad [6.14]$$

where:

$$\begin{aligned} S_0 &= E_{x0}^2 + E_{y0}^2 \\ S_1 &= E_{x0}^2 - E_{y0}^2 \\ S_2 &= 2E_{x0}E_{y0}\cos\delta \\ S_3 &= 2E_{x0}E_{y0}\sin\delta \end{aligned} \qquad [6.15]$$

The parameters S_0, S_1, S_2 and S_3 are known as the *Stokes polarization parameters* for a plane wave, and were introduced in 1852 by Sir George Gabriel Stokes.[10] They provide a phenomenological description of the state of polarization of light, involving only the optical intensity, which can be experimentally determined.

The first parameter expresses the total intensity of the optical field while the remaining three parameters describe the polarization state of light. The Stokes vector helps in the measurement of:

1. the intensity of light;
2. the DOLP with respect to vertical and horizontal axes;
3. the DOLP with respect to the axes oriented at +45° and −45°;
4. the degree of left- and right-circular polarizations.

One of the main advantages of the Stokes formalism is that it allows for the determination of the degree of polarization (DOP) of any optical signal directly from its Stokes components. In most cases, an electromagnetic wave consists of the incoherent superposition of a fully polarized wave and of a fully unpolarized wave. The DOP quantifies that fraction of the total optical signal that is actually polarized, with:

$$\mathrm{DOP} = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0} \qquad [6.16]$$

The DOP of a fully polarized optical signal is equal to 1, whereas the DOP of an unpolarized optical signal is 0. Circular and elliptical polarizations do not usually occur in a linearly polarized light beam. Its DOP is thus often referred to as the degree of linear polarization. The DOLP of a light beam is defined by:

$$\mathrm{DOLP} = \frac{S_1}{S_0} \qquad [6.17]$$

The Stokes DOLP is thus a ratio of the difference between horizontal and vertical intensity components and the total normalized light intensity.

6.3 Polarization cameras

The polarization of light can be easily visualized using polarization filters. A polarization filter has the property either to transmit light vibrating in one direction producing linearly polarized light, or light with a sense of rotation producing circularly polarized light. In essence, a polarization filter translates polarization information into intensity difference information. This is important as humans are able to perceive intensity information but not polarization information.

A polarization camera is a generalization of the conventional intensity camera. To obtain parameters describing the polarization state of monochromatic light, a set of measurements has to be taken from the different positions of the polarizers, waveplates and photodetectors. Therefore, the duration of data acquisition time depends on how fast the polarizers, waveplates and photodetectors are moved to their various measurement positions. In addition, the precision of positioning the polarizers, waveplates and photodetectors influences the results of measurements.

There are three different types of polarization cameras available: (1) cameras with an external polarization filter, (2) cameras with organic polarization filters and (3) cameras with wire-grid polarization filters. In this section, these three types of polarization camera are discussed.

6.3.1 Polarization filters types

State-of-the-art polarization image sensors consist of a photodetector array, such as a (CMOS) or a CCD sensor array, and a single or multi-axis micropolarizer array to measure the polarization information in real time. Based on the position of the micropolarizers and the mode of data acquisition, the current state-of-the-art polarization image sensors can be divided into two types: time domain multiplexed mode sensors, and spatial domain multiplexed mode sensors.

The first group of polarization sensors includes standard imaging sensors coupled with electrically or mechanically controlled polarization filters.[11] These are polarization cameras with external polarization filters. They are operated in time domain multiplexed mode: the externally coupled linear polarizer is rotated in front of either a single photodetector or a complete photodetector array. The main disadvantage of these systems is that a mechanical or electrical rotation of the polarizer filter is required, thus it is not suitable for real time designs. Multiple images with a different orientation of the polarization filter are required for the measurement of partial linear polarization, increasing the frame time by the desired total number of orientation measurements. Furthermore, as the precision of the position of the polarizers affects the measurement results, these systems often produce erroneous results. However, these systems have high spatial resolution compared with organic and wire-grid micropolarizers, as the whole imaging array images only one orientation at a time. Commercially available external linear polarizers include dichroic glass polarizers from Polarcor™, Thorlabs and Edmund optics. These glass polarizers usually use proprietary stretched polyvinyl alcohol (PVA) films to create the polarization axis and exhibit extinction ratios greater than 10 000.

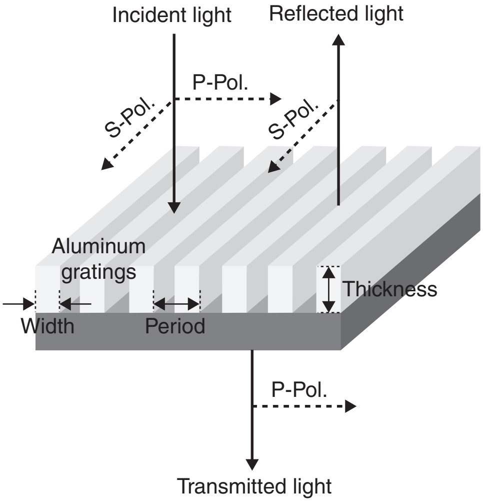
The second group of polarization sensors includes image sensors where the linear micropolarizer is either fabricated directly on the sensor array, or on a substrate and which is then flip-chip bonded with the sensor array. Monolithic integration of micropolarization filters with CMOS imaging sensors is currently a subject of intense research. A plane coated with micrometer-scale polarizing elements in a variety of orientations is called a micropolarizer array. They operate in spatial domain multiplexed mode, in which the linear polarizers are oriented in different directions with respect to the photodiodes and are directly fabricated on top of the photodiode. Micropolarizers have been fabricated with metal grids and organic materials. Incorporating pixel-pitch-matched polarization filters at the focal plane for the visible spectrum has been explored with birefringent materials[3] and thin film polarizers.[12,13]

Photolithographically patterning polymer polarization film to make micropolarizer arrays was proposed by Faris.[14] Polymer film micropolarizer arrays have been successfully deployed in display applications, but their use in imaging is more challenging. For imaging, the thickness of the polarizing element has to be much smaller than the pixel size in order to have a large light collection angle. Polymer polarizers are made by mechanically stretching the polymer film. It is difficult to manufacture uniform stretched films with a thickness of less than 10 μm.[15] Another main disadvantage of the thin film polymer polarization imaging sensors has been the complex fabrication techniques required to merge CMOS technology with polymer filters. These fabrication steps are prone to misalignment errors, which can degrade the sensing capabilities of the polarization imaging sensor. Furthermore, the thick multilayer filter array, which is around 10 μm of thickness, is prone to optical crosstalk and limits the extinction ratios of the sensors.[16] The spatial orientation of the linear polarizer filter results in reduced spatial resolution. However, since the linear polarizers are directly incorporated on top of the photodiode, the temporal resolution is increased – this allows for faster polarization imaging and also contributes to the compactness of the design. The precision of the system also increases, as the position of the polarizer is fixed.

Micropolarizers with wire-grids are planar structures composed of a series of parallel wires or strips made of a conducting material. Wire-grid polarizers are inorganic optical polarization devices made of aluminum. Aluminum is usually chosen for the wire-grid polarizer in the visible region of the electromagnetic spectrum as it is highly reflective and absorbs only a small portion of the incident electromagnetic wave in the visible region. Furthermore, they are able to withstand high temperatures, while the performance of organic polarizers is typically degraded at high temperatures.[17] For example, organic polarizers made of stretched PVA start degrading at temperatures near 150°C, while aluminum wire-grid polarizers can withstand temperatures above 350°C. The aluminum wire-grid polarizer also provides high transmission efficiency and a polarization extinction ratio. Thus polarization cameras with wire-grid polarizers are an ideal choice for machine vision applications. Commercially, Edmund optics and NIL Technology produce wire-grid polarizers. Edmund optics makes the wire-grid polarizers using thin layers of aluminum MicroWires™ between two glass windows while NIL uses metallic nano-wires.

Usually, a linear wire-grid would function as a simple diffraction grating but, when the pitch or period of the wires is less than half the wavelength of the incoming light, it becomes a polarizer. Wire-grid polarizers are extensively used as polarization-sensitive elements in various applications because of their excellent polarization performance and planar structure that allows them to be easily pixilated and integrated into other optoelectronics devices. A schematic view of a wire-grid polarizer is shown in Fig. 6.7.

Metal grid micropolarizer arrays are often used for infrared imaging polarimetry,[18] but they are not very effective in the visible wavelength range. For the wire-grid polarizer to be effective in the visible region, for optimum performance the pitch of the grid needs to be at least three times smaller than the smallest wavelength that is being polarized. Thus, for the efficient polarization of a spectrum with a minimum wavelength of 400 nm, the wire-grid polarizer should have a pitch of, at most, 133 nm. With the scaling of the modern CMOS processes, the metal planes normally used for interconnections when specifically aligned are now becoming suitable for polarization transmission.[19]



6.7 Wire-grid polarizer.

Table 6.1 Comparison of linear polarizers

	External filter	Organic micropolarizer	Metallic wire-grid polarizer
Operating mode	Time domain	Spatial domain	Spatial domain
External control	Required	Not required	Not required
Extinction ratio	Very high	High	Low
Spatial resolution	High	Low	Low
Temporal resolution	Low	High	High
Additional fabrication required?	No (externally coupled)	Yes	No (standard CMOS metal layers are used)
Commercial availability	Polaroid, Polarcor™, Thorlabs, Edmund optics	Polarcor™, Thorlabs, Edmund optics	Edmund optics, NIL Technology. CMOS metallic grid is still in research and not commercially available

Table 6.1 summarizes the discussion on different types of linear micropolarizers. The extinction ratio is the ratio of the amount of light transmitted to the amount of light blocked and is discussed in detail in Section 6.3.2.

Wire-grid polarizers are extensively used as polarization sensitive elements in various applications because of their excellent polarization performance and planar structure that allows them to be easily pixilated and integrated into other optoelectronics devices. Thus polarization cameras with wire-grid polarizers are an ideal choice for machine vision applications.

6.3.2 Polarization characterization and measurements

When randomly polarized light is transmitted through an ideal wire-grid polarizer, the electrical fields orthogonal to the wires will be transmitted and electric fields parallel to the wire will be reflected (as shown in Fig. 6.7). In practice, however, wire-grid polarizers transmit some undesired parallel electrical fields while reflecting a desired orthogonal electrical field and also absorbing some of the incident light. Thus, the transmitted electrical field through a wire-grid polarizer can be modeled as having two components: an electrical vector associated with the parallel electric field, and an electric vector associated with the orthogonal electric field – each scaled by a scalar value. The scalar value is proportional to the square root of the intensity associated with each polarization. Equation [6.18] is the resulting electric vector equation for unpolarized light passing through a wire-grid polarizer.

$$\vec{E} = t(\hat{p}\cdot\vec{E})\hat{p} + c(\hat{r}\cdot\vec{E})\hat{r} \qquad [6.18]$$

where E is the resultant electrical field after transmission through the wire-grid polarizer and the scalar value for the parallel electric field $(\hat{p})$ is t and the scalar value for the orthogonal electric field $(\hat{r})$ is c.

The irradiance of the plane polarized light passing through a polarizer is given by the Malus's law.[20] If the transmission axis of the polarization filter is parallel to the incident polarized light, the transmitted intensity is given by:

$$I = I_0 \cos^2 \varphi \qquad [6.19]$$

where I is the transmitted intensity, I_0 is the maximum transmitted intensity and φ is the angle between the polarizer's transmission axis and the plane of polarization of the incident light. Thus, the maximum intensity that can be sensed by the detector is dependent on the angle of the transmission axis of the polarization filter and would result in a cosine wave with a varying transmission axis angle.

Based on the transmission of intensity at a given transmission axis of a linear polarizer, linear polarizers are characterized by two main specifications: transmittance efficiency, and extinction or contrast ratio. The wire-grid polarizer's *transmittance efficiency* is defined by the fraction of the total incident light that is transmitted through the linear polarizer. The maximum transmission in the linear polarizer is obtained when the polarizer orientation is aligned parallel to the orientation of the linear polarized component; minimum transmission is obtained when the orientation of the polarizer is aligned perpendicular to the linear polarized component of light.

The extinction ratio (ER) is a measure of the polarization contrast of a linear polarizer and it is defined as the intensity ratio between the maximum and minimum throughputs of the polarizer when the wire-grid is aligned parallel and perpendicular with respect to the light beam. The accuracy of the wire-grid polarizer's extinction ratio measurement depends on several factors – such as the detector precision, analyzer alignment, stray light variations and light source stability. In an ideal testing environment with no light source or detector drift or background noise variations, the extinction ratio can be simply defined as:

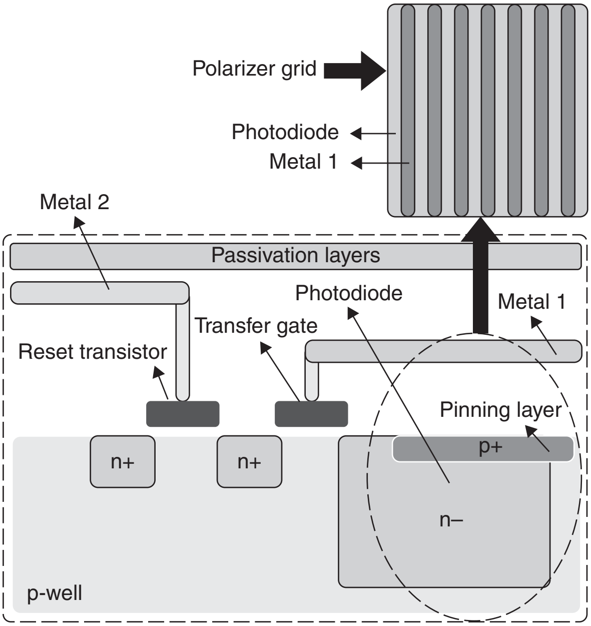
$$\mathrm{ER} = \frac{I_p}{I_s} \qquad [6.20]$$

where I_p is the maximum transmittance parallel to the plane of the polarized beam and I_s is the minimum transmittance perpendicular to the plane of the polarized beam. The contrast ratio for polarizers varies depending on the polarizer, ranging from around 10 to over 10^7.

6.4 Design of a polarization sensor

In Section 6.3.1, it was shown that, in order to absorb the electromagnetic waves in the visible spectrum selectively, the grid pitch should be very small. With the scaling of the technology nodes, is it now possible to obtain a smaller grid pitch with a standard CMOS process. In a 0.18 μm standard CMOS process, the minimum allowable distance between two metals is 0.24 μm. A one-dimensional wire-grid polarizer implemented using the bottom metal layer (M1) is shown in Fig. 6.8.

The wire-grid polarizer is implemented over a pinned photodiode ($p^+/n^-/p$-sub), described in Sarkar.[7] The thickness of the metal layer M1 used to create the wire-grid polarizer is 0.35 μm, while the width of the metal wire is 0.24 μm. The choice of the thickness and the width of the metal wire are limited by the choice of the technology. The aspect ratio (thickness versus width) for the designed wire-grid micropolarizer is 1.5:1. Based on the selected grid width and height, the incident *p*-polarized light (the electrical


Polarizer grid
Photodiode
Metal 1
Metal 2
Passivation layers
Photodiode
Metal 1
Transfer gate
Reset transistor
Pinning layer
p+
n+
n+
n–
p-well

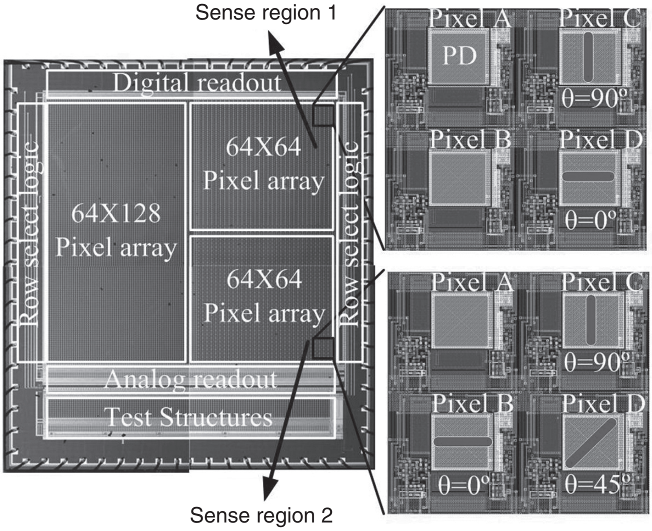
6.8 Pixel cross-section and top view of the implemented wire-grid polarizer.

field of which is parallel to the grating) is mostly blocked and the *s*-polarized (the electrical field of which is perpendicular to the grating) is mostly transmitted and detected by the photodiode.

Figure 6.9 shows the implementation of the polarization analyzing CMOS image sensor operated in spatial mode. The wire-grid micropolarizer was created with varying orientations over the pixel array to enable real time measurement of the polarization information. The active pixel array of 128 × 128 pixels was split into two regions of 64 × 128 pixels. One of these regions was used for normal imaging applications with the photodiode free of wire-grid micropolarizer; the other region was made polarization sensitive by using the wire-grid micropolarizer on top of the photodiode. The region selected to analyze the polarization was further split into two regions of 64 × 64 arrays, referred to as sense regions 1 and 2. A 2 × 2 pixel array in sense region 1 consists of two pixels (A and B) measuring the intensity, while the other two pixels measure, individually, the 0° (D), and 90° (C) degree polarized intensity. A 2 × 2 pixel array in sense region 2 has one pixel recording the intensity of the light (A) while the other three record the 0° (B), 45° (C) and 90° (D) polarized intensity.

In this chapter, only polarization region 1 is considered for machine vision applications. The pixels for normal imaging and region 2 are used for differential imaging and polarized based navigation, respectively, and are not discussed in this chapter.

The polarizer in region 1 was fully characterized. For 0° polarization, the maximum transmittance achieved was 38.9% and the minimum


Sense region 1
Digital readout
Row select logic
64X128
Pixel array
64X64
Pixel array
64X64
Pixel array
Row select logic
Analog readout
Test Structures
Sense region 2
Pixel A
PD
Pixel C
θ=90°
Pixel B
Pixel D
θ=0°
Pixel A
Pixel C
θ=90°
Pixel B
θ=0°
Pixel D
θ=45°

6.9 Sensor regions with different polarizing angles.

transmittance observed was 7%. For 90° polarization, the maximum transmittance obtained was 44.4% and the minimum transmittance was 0.1%. The calculated extinction ratio for the linear polarizer in the sense region 1 was 6.3.[7] The extinction ratio was limited by the fabrication technology used; with lower technology nodes, a higher extinction ratio can be obtained.

As noted previously, commercial organic polarizers or wire-grid polarizers have an extinction ratio of over 100, the extinction ratio obtained with a conventional CMOS metal grid is somewhat lower. The polarization image sensors with conventional CMOS metal grida would thus not be suitable for polarization imaging. Instead, these sensors can be used for applications where high extinction ratios are not required. One such application is material classification. The specular or diffused reflection from a material surface is polarized; the DOP varies with the nature of the material surface as well as the direction of the incoming beam of light. The DOP of the reflected polarized light for a metal and a dielectric surface would be considerable and thus would not need a high extinction ratio for differentiating between the two surfaces. Thus, the CMOS image sensor designed for the detection of polarization with an embedded metallic wire-grid polarizer can be used for real time material classification, the theory of which and the relevant experiments are presented in Section 6.5.

6.5 Polarization vision in machine vision applications: material classification

Material classification is an important application in computer vision. The ability to detect the nature of the object surface from image data has considerable potential for applications ranging from low-level inspection to high-level object recognition. The inherent property of materials partially to polarize the reflected light can serve as a tool with which to classify them.

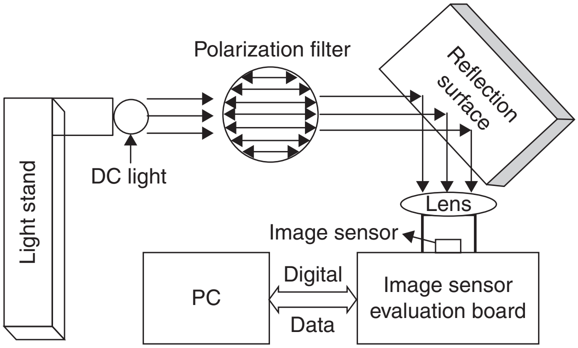
Materials can be broadly classified as either metals or dielectrics, based on their conductivity. Metals are highly conductive and opaque, and tend to be very reflective; dielectrics are less conductive and have very low reflectivity. Initially, the dichromatic reflection properties of the material surface[21–23] were used to distinguish between metals and dielectrics. Materials were classified as either optically homogenous or optically inhomogeneous. Homogenous materials reflect light from the surface only;[24] thus, when illuminated with a monochromatic light beam, the reflected light color would be almost constant. Inhomogeneous materials, on the other hand, reflect light from the surface and also scatter light from the body. Such a material, when illuminated with a monochromatic light beam, would reflect two distinct colors as the color of the surface reflected light would be different from the color of the scattered light from the body. This method, however, is limited by the distortion of the color histogram obtained after reflection, which depends on the object geometry. Also, not only the color of the material, but also the color of the light influences the reflected color. Another method with which to discriminate between metals and dielectrics based on the Fresnel reflection theory was proposed by Wolff.[2,25] According to Fresnel reflection theory,[2] dielectric surfaces strongly polarize the light upon specular reflection for all angles of specular incidence when compared with the reflected light from metal surfaces.

Traditionally the polarization components of the transmitted irradiances are obtained by allowing the reflected light from the material surface to pass through an external linear polarizer onto a CCD or CMOS image sensor. When the reflected wave is passed through a linear polarizer, the intensity of the image, $I(x,y)$ can be expressed as a function of the transmission axis of the polarizer $I(x,y,\theta)$, where θ is the transmission axis of the linear polarizer. The reflected irradiance oscillates between the maximum transmitted intensity $I_{\max}$ and the minimum transmitted intensity $I_{\min}$. The magnitude of oscillation is quite large for dielectrics as compared with metals.[2] The disadvantage of such a system is that the linear polarization filters have to be externally controlled, which complicates the automation and miniaturization of optical sensors for material classification. Further, the digital processing required to process the obtained polarization information increases the complexity and power consumption of the system.

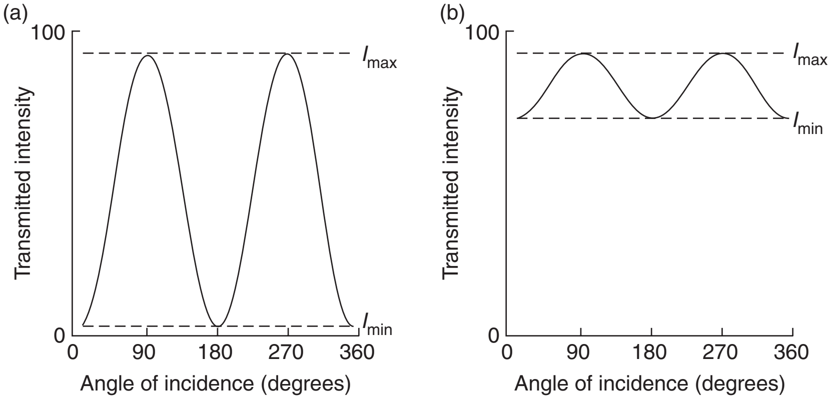
The embedded wire-grid metallic micropolarizer can be used for material classification instead. The metallic wire-grid micropolarizer is used to measure the transmitted irradiances after specular reflection from the material surface in real time. The metal grid oriented at 0° and 90°, patterned over the photodiodes, allows for the measuring of the parallel and perpendicular Fresnel reflection coefficients and the maximum and the minimum transmitted irradiance following reflection from the material surface.[26]

The measurement set-up for the measuring of the maximum and minimum transmitted intensities following reflection from the material surface is shown in Fig. 6.10. The light from the DC source passes through a linear polarizer. The transmission axis of the linear polarizer was varied from 0° to 180° in steps of 15° to vary the polarization angle of the incoming DC light. The polarized electromagnetic waves were then reflected from the metal and dielectric surface. At the boundary of the reflection surface, both the diffused component and the specular component of the reflection of the incident light were present. These reflection components were then incident on the image sensor after having been focused by a lens. The analogue signal from the image sensor was digitized using an external ADC and then analyzed using a PC.

For dielectrics, the Fresnel coefficients satisfy $R_p >> R_s$ while for metals $R_p \approx R_s$, where R_P and R_S are the parallel and perpendicular Fresnel reflection


Polarization filter
Reflection surface
Light stand
DC light
Lens
Image sensor
PC
Digital
Data
Image sensor evaluation board

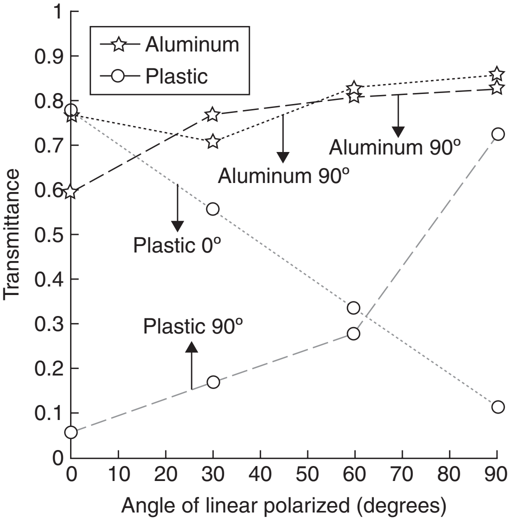
6.10 Measurement set-up.



6.11 Transmitted radiance of reflected light as a function of transmission axis of the linear polarizer: (a) dielectric surface, (b) metal surface.

coefficients, respectively. The diffused component of reflection is greater than the specular component for dielectrics; thus, the oscillations between I_{max} and I_{min} vary over a larger range. In case of metals, the oscillations are relatively smaller as the specular component of reflection is greater than the diffuse component of reflection. Thus, as noted previously, the magnitude of oscillations of the reflected irradiance of the light after reflection is greater for dielectrics than for metals.[2,27] The theoretically expected image intensity $I(x,y,\theta)$ of the transmitted radiance of light reflected from the dielectric and metal surfaces is shown in Fig. 6.11.[27]

Figure 6.12 shows the measured transmitted irradiance for the pixels with patterned 0° and 90° metallic wire-grid, with the linear wire-grid polarizer implemented using thin metal strips (as shown in Fig. 6.9).[26]



6.12 Transmitted intensity at 0° and 90° polarization-sensitive pixel.

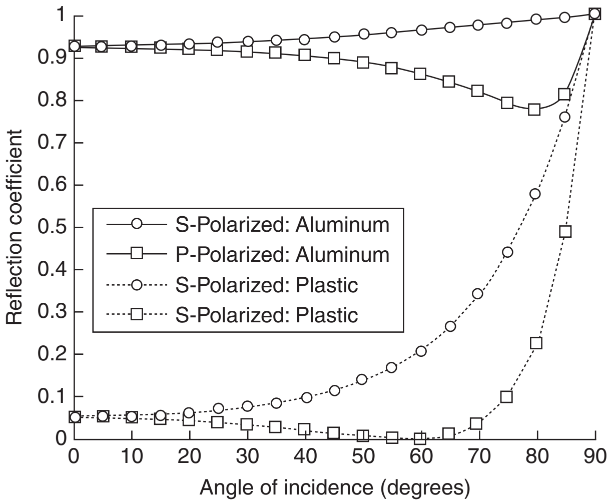
Table 6.2 Transmitted radiance for plastic and aluminum

Transmitted intensity	Plastic	Aluminum
$I_{max} - I_{min}$ (0°)	0.66	0.15
$I_{max} - I_{min}$ (90°)	0.67	0.23

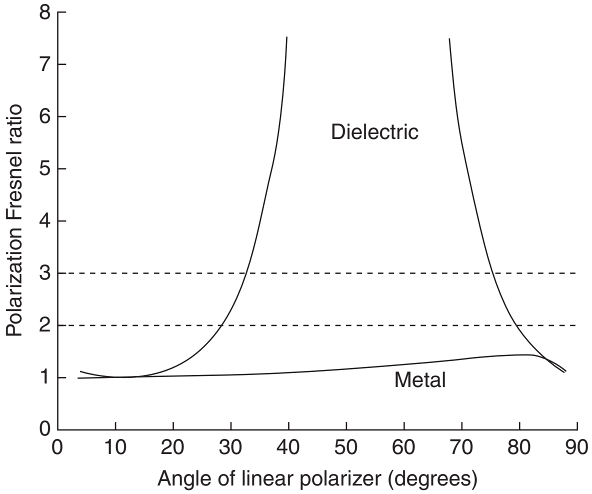
The difference between normalized maximum and the minimum transmitted irradiance for plastic and aluminum surfaces is shown in Table 6.2. The transmitted irradiance oscillations for plastic are 190% and 340% higher than those for aluminum. The differences in the transmitted irradiance are due to the difference in the reflection pattern of the light from the aluminum and plastic surfaces.

Besides the transmitted intensities, the polarization Fresnel ratio (PFR) also serves to distinguish between metals and dielectrics. The PFR is the ratio of the perpendicular Fresnel coefficient to the parallel Fresnel coefficient. The theoretical Fresnel reflection coefficients for aluminum and plastic are shown in Fig. 6.13. The Fresnel reflection coefficient for aluminum is near 1, while that for plastic varies over the entire span from 0 to 1 for different specular angles of incidence.

A material with significant conductivity will have a significantly reduced PFR over a large range of specular angles of incidence. Since the conductivity of metals is higher than the conductivity of dielectrics, the PFR for metals is much smaller compared with that of dielectrics. As in metals, the

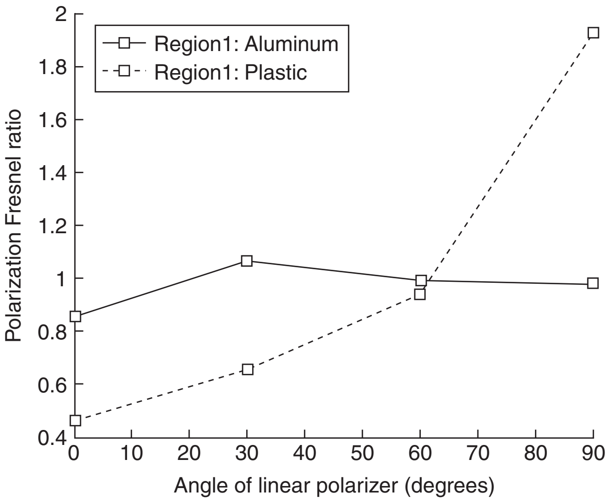


6.13 Theoretical Fresnel reflection coefficients for aluminum and plastic at $\lambda = 500$ nm.[2]



6.14 Theoretical polarization Fresnel ratio.[2]

specular component of reflection is greater than the diffused component of the reflection, the Fresnel coefficients satisfy $R_s(x,y) \approx R_p(x,y)$.[27] Using this condition, it can be said that the PFR of the metals for which $I_{\max}(x,y) \approx I_{\min}(x,y)$ will almost be equal to 1 for all angles of incidence. The theoretical PFR for metal and dielectric surfaces is shown in Fig. 6.14.[2]

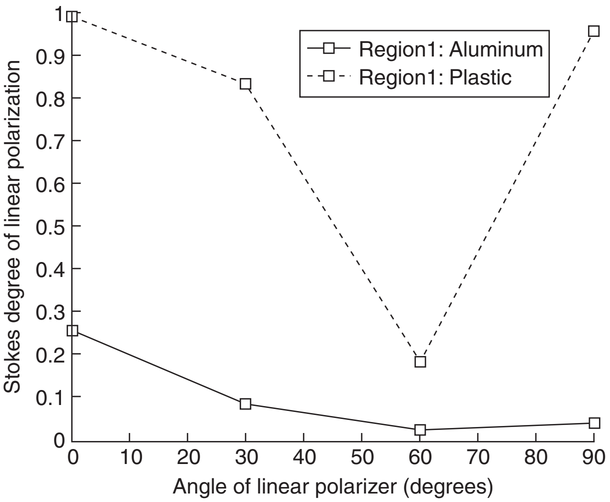


6.15 Polarization Fresnel ratio of 0° and 90° polarization-sensitive pixel.

The $I_{\max}(x,y)$ and $I_{\min}(x,y)$ values are determined using the same measurement set-up shown in Fig. 6.10. For varying transmission axis of the external linear polarizer, the pixel outputs averaged over 30 frames at the 0° and 90° polarization-sensitive pixels outputs are stored. The transmittance is computed by normalizing the output at 0° and 90° by the intensity measured at the pixels with no metallic wire-grid. The PFR is then calculated from the maximum and minimum transmittance. The experimentally obtained PFRs for aluminum and plastic are shown in Fig. 6.15.

Figure 6.15 shows a clear threshold in the PFR values for metals and dielectrics. The PFR for aluminum is usually in the range of 0.8–1 for all specular angles of incidence, while the PFR for dielectrics rapidly increases near the Brewster angle.[28] At the Brewster angle, the reflected light is completely polarized perpendicular to the plane of the incidence. Theoretically, for a specular angle of incidence greater than the Brewster angle, the PFR for aluminum is always smaller than 2.[2] A PFR value of almost 2 can be considered to belong to a dielectric. The PFR computations offer a good match to the theoretical studies and thus can be used to classify the materials as either metal or dielectric.

As discussed in Section 6.2.3, the polarization state of an electromagnetic wave can be conveniently described by a set of Stokes parameters. The Stokes parameters have not been previously used for classifying material. The Stokes parameters S_0 and S_1 are calculated from the 0° and 90° degree



6.16 Stokes DOLP for 0° and 90° polarization-sensitive pixel.

polarization-sensitive pixels; from these, the Stokes DOLP is obtained using Equation [6.17] (shown in Fig. 6.16).

The Stokes DOLP for plastic has a maximum value of 1 and is higher than that for aluminum. It is further observed that the Stokes DOLP for plastics steadily decreases until it reaches the Brewster angle, where the DOP again increases to its maximum value. The maximum Stokes DOLP for aluminum obtained is 0.25.[28] The DOLP for aluminum is low compared with that for plastic for all specular angles of incidence. Furthermore, the DOLP of plastic is found to decrease with the increase of the specular angle of incidence and a sharp rise in the DOLP is observed around the Brewster angle for plastics.

6.6 Conclusion

Lighting for machine vision is associated with special requirements. The nature of the light source used and the surface properties of the object being imaged determine the contrast perceived by the vision systems. Specular objects are difficult to image since they reflect the entire incident light, saturating the vision system and also producing glare. Polarization imaging is helpful in reducing the glare from the reflections of the specular surface. In polarization imaging, a polarized light is incident on the object and reflected light with the same polarization state is rejected by an analyzer placed in front of the vision system while allowing other diffused light to pass through. This helps in reducing the glare as well as the transmitted intensity, preventing saturation of the vision system.

The property of the object to polarize the light incident on it can be used for real time material classification. The state of polarization for the diffused and specular components of the reflection depends on the reflecting surface. The measurement of the state of polarization of the reflected light serves as an indicator for the type of material surface. The magnitude of oscillations of the maximum and minimum transmitted irradiance due to the variation in the reflection pattern can be used to distinguish between a metallic surface and a dielectric surface.

Besides the transmitted intensities, the PFR and the Stokes degree of linear polarization are also useful in distinguishing between metal and dielectric surfaces. The ratio of the irradiance measured at 0° and 90° polarization-sensitive pixels can be obtained using an analogue divider, which then relates to the PFR. The PFR for metals does not vary much with the variations in the linear polarizer angle and thus metals are easily distinguishable from dielectrics, which show marked variance.

Polarization imaging helps in automation of the measurements/detection of material surface in real time and thus can serve as an efficient tool in machine vision applications. The polarization profile of the reflected light, for example, depends on the surface characteristics and thus can be used for their intelligent classification into conductive/non-conductive, or to determine the smoothness/roughness of a surface. Thus, polarization imaging has the potential to become a powerful tool to enhance information detection and signal processing in industrial applications.

6.7 References

1. D. Martin, 'A practical guide to machine vision lighting', *Advanced Illumination,* pp. 1–18 (2007).
2. L. Wolff, 'Polarization based material classification from specular reflection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. **12**, no. 11, pp. 1059–1071 (1990).
3. A.G. Andreou and Z.K. Kalayjian, 'Polarization imaging: principles and integrated polarimeters', *IEEE Sensors Journal*, vol. **2**, no. 6, pp. 566–576 (2002).
4. G.D. Bernard and R. Wehner, 'Functional similarities between polarization vision and color vision', *Vision Research*, vol. **17**, pp. 1019–1028 (1977).
5. R. Wehner, 'Polarization vision – a uniform sensory capacity?', *Journal of Experimental Biology*, vol. **204**, pp. 2589–2596 (2001).
6. D. Reiley, 'Polarization in optical design', PhD Dissertation, University of Alabama, pp. 2–4 (1993).
7. M. Sarkar, 'A biologically inspired CMOS image sensor', PhD Dissertation, Technical University of Delft, pp. 31–100 (2011).

8. G. Horváth and D. Varjú, '*Polarized Light in Animal Vision*', Berlin, Heidelberg/New York, Springer, ISBN: 3540404570 (2003).
9. http://www.microscopyu.com/articles/polarized/index.html.
10. G.G. Stokes, *Transactions of Cambridge Philosophical Society*, vol. **9**, p. 399 (1852). Reprinted in *Mathematics and Physics Papers*, vol. 3, Cambridge University Press, pp. 203–233 (1901).
11. J. Tyo, D. Goldstein, D. Chenault and J. Shaw, 'A review of passive imaging polarimetery for remote sensing applications', *Applied Optics*, vol. **45**, pp. 5453–5469 (2006).
12. X. Zhao, F. Boussaid, A. Bermak and V. Chigrinov, 'Thin photopatterned micropolarizer array for CMOS image sensors', *IEEE Photonics Technology Letters*, vol. **21**, no. 12, pp. 805–807 (2009).
13. V. Gruev, J.V. der Spiegel and N. Engheta, 'Image sensor with focal plane polarization sensitivity', *Proceedings of International Symposium on Circuits and Systems*, pp. 1028–1031 (2008).
14. S. Faris, 'Methods for manufacturing micropolarizers, US patent 5,327,285 (1994).
15. J. Guo and D. Brady, 'Fabrication of thin-film micropolarizer arrays for visible imaging polarimetry', *Applied Optics*, vol. **39**, no. 10, pp. 1486–1492 (2000).
16. V. Gruev, A. Ortu, N. Lazarus, J.V. der Spiegel and N. Engheta, 'Fabrication of a dual-tier thin film micropolarization array', *Optical Express*, vol. **15**, no. 8, pp. 4994–5007 (2007).
17. J.J. Wang, F. Walters, X. Liu, P. Sciortino and X. Deng, 'High-performance, large area, deep ultraviolet to infrared polarizers based on 40 nm line/78 nm space nanowire-grids', *Applied Physics Letters*, vol. **90**, pp. 061104-1-3 (2007).
18. G. Nordin, J. Meier, P. Deguzman and M. Jones, 'Micropolarizer array for infrared imaging polarimetry', *Journal of Optical Society of America*, *A*, vol. **16**, pp. 1168–1174 (1999).
19. T. Tokuda, H. Yamada, K. Sasagawa and J. Ohta, 'Polarization analyzing CMOS image sensor with monolithically embedded polarizer for microchemistry systems', *IEEE Transactions on Biomedical Circuits and Systems*, vol. **3**, no. 5, pp. 259–266 (2009).
20. E. Collett, '*Field Guide to Polarization*', Bellingham, SPIE Press, WA, p. 25, ISBN: 9780819458681 (2005).
21. G. Klinker, S. Shafer and T. Kanade, 'Using a color reflection model to separate highlights from object color', Proceedings of International Conference on Computer Vision, pp. 145–150 (1987).
22. S. Shafer, 'Using color to separate reflection components', *Color Research Applications*, vol. **10**, no. 4, pp. 210–218 (1985).
23. G. Healey and W.E. Blanz, 'Identifying metal surfaces in color images', *Proceedings of Conference in Optics, Electro-Optics, and Sensors* (1988).
24. G. Healey and T.O. Binford, 'Predicting material classes', *Proceedings of DARPA Image Understanding Workshop*, pp. 1140–1146 (1988).
25. H. Chen and L.B. Wolff, 'Polarization phase based method for material classification and object recognition in computer vision', *International Journal of Computer Vision*, vol. **28**, no. 1, pp. 45–56 (1999).

26. M. Sarkar, D. San Segundo, C. van Hoof and A.J.P. Theuwissen, 'Integrated polarization analyzing CMOS image sensor', *Proceedings of IEEE International Symposium on Circuits and Systems*, pp. 621–624 (2010).
27. S. Tominaga and A. Kimachi, 'Polarization imaging for material classification', *Optical Engineering*, vol. **47**, no. 12, pp. 123201-1-14 (2008).
28. M. Sarkar, D. San Segundo, C. Van Hoof and A.J.P. Theuwissen, 'Integrated polarization analyzing CMOS image sensor for real time material classification', *IEEE Sensors Journal*, vol. **11**, pp. 1692–1703 (2011).

7

Advanced interfaces for resistive sensors

A. FLAMMINI and A. DEPARI,
University of Brescia, Italy

DOI: 10.1533/9780857099297.1.153

Abstract: Resistive sensors are among the most widespread devices in data acquisition systems. This is mainly due to the broad availability of sensors for different applications. Due to the simplicity of the sensing devices and the opportunity to implement simple and low-cost data readout systems, they are of considerable value when the focus is on the realization of smart sensors. In this chapter, traditional resistive sensor interface methods will be analysed; more advanced techniques appropriate for specific applications, as well as targeting the effectiveness of the acquisition systems, will subsequently be explored. The chapter closes with discussion regarding future trends.

Key words: resistive sensor interface, wide-range resistance measurement, resistance-to-time conversion method, parasitic capacitive effects, weighted least mean squares linearization.

7.1 Introduction

The success of resistive sensors is mainly due to their availability for many different applications: thermal and light detectors, gas presence and concentration evaluation, position measurement, strain estimation to name just a few examples of the possible fields of application. The simplicity of the sensing devices together with the opportunity to implement simple and low-cost sensor interface circuits make this kind of device highly appropriate for the realization of smart sensors – in which the sensing element, the signal conditioning electronics and the data acquisition/elaboration/transmission stages form a unique unit.

The chapter is structured as follows: Section 7.2 gives a brief introduction and general information about resistive sensors; Section 7.3 details the electronic interface perspective, starting with an analysis of the traditional interface methods. In this section, calibration and linearization procedures are illustrated. Section 7.4 explores more advanced techniques that are suitable for specific applications (e.g., wide-range sensors), or for enhancing the effectiveness of the acquisition systems (e.g., coping with parasitic capacitive effects). Finally, Section 7.5 presents a discussion related to future trends, and conclusions are offered.

7.2 Resistive sensors

A resistive sensor provides information in the form of the electrical resistivity of a material or electrical resistance of a device.

The electrical resistivity ρ is the opposition of a specific material to the flow of an electric current when an electrical field is applied; it depends on the properties of the material itself. The reciprocal of the resistivity is the conductivity σ. The alteration of the physical/chemical conditions of the resistive sensor, caused by the physical quantity of interest, results in variations in its resistivity. However, being a microscopic effect, the estimation of electrical resistivity is quite complicated and not achievable with tools and techniques available for the realization of smart sensors.

The electrical resistance R is the opposition of a device to the flow of an electric current when an electrical potential is applied. The reciprocal of the resistance is the conductance G. Since it is a property related to an object, made of a specific material and with a determined shape and size, the electrical resistance is a macroscopic effect and thus it is easier to estimate using different approaches.

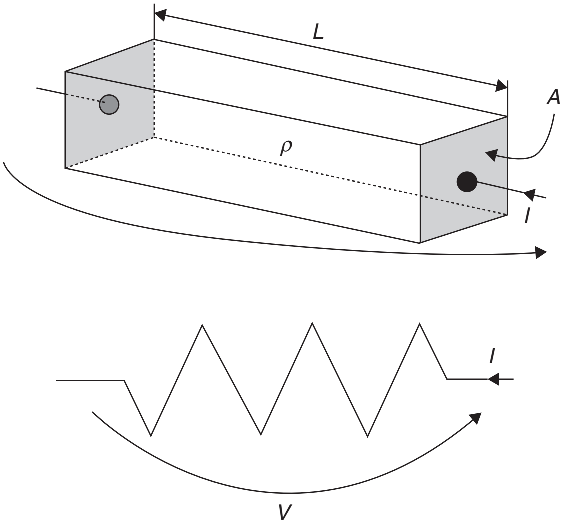
Ohm's Law (Equation [7.1]) furnishes the relationship between the applied voltage V and the current I flowing through a device, and gives the basic concept regarding resistance estimation techniques, which will be shown in the subsequent sections.

$$V = RI \tag{7.1}$$

If a device is made of a homogeneous material, the shape and the size determine the relationship between the electrical resistivity of the material and the electrical resistance of the object. For example, given a parallelepiped with cross-section A and length L (as in Fig. 7.1, where the conventional symbol of the electrical resistance is reported as well), Equation [7.2] shows how to relate the resistance R (conductance G) to the resistivity ρ (conductivity σ).

$$R = \rho\frac{A}{L}; \quad G = \sigma\frac{A}{L} \tag{7.2}$$

The measurement of the electrical resistance is therefore an indirect way of measuring the electrical resistivity. In addition, it should be noted that,



7.1 The parallelepiped shape usually employed to relate the electrical resistivity to the resistance and the conventional symbol used to identify the electrical resistance.

even if the geometrical properties of the object (the ratio L/A in Equation [7.2]) are unknown or are not easy to compute (e.g., irregular object shape), the direct proportionality between R and ρ shown in Equation [7.2] is still valid. This leads to a direct proportionality between resistivity variation $\Delta\rho$ and resistance variation ΔR.

7.2.1 Examples of resistive sensors

Examining Equation [7.2], it should be noted that the resistance value R is the combination of a physical factor (ρ) and a geometrical factor (the ratio L/A). Resistive sensors can convert variations in any such parameter a physical quantity of interest into a variation of resistance. The five main types of resistive sensor are discussed below; however, this selection is not exhaustive.

Resistive temperature detectors (RTDs)

As the name suggests, these sensors relate resistance to the environmental temperature. They are usually made of a metallic filament wrapped in a ceramic or other dielectric material. The basic working principle is based on the dependence on temperature of electron mobility in the metallic filament; thus, this kind of sensor employs the variation of the resistivity ρ. It should be pointed out that the size of the filament can also change according to temperature, but the effect of the geometrical variations on sensor resistance is negligible with respect to the resistivity variation.

When a conductive material is employed, sensor resistance usually increases with temperature (positive temperature coefficient – PTC); alternative techniques and materials, such as semiconductors, can be used as the sensing component, demonstrating the opposite effect: reduction of the sensor resistance with increasing temperature (negative temperature coefficient – NTC).

Light dependent resistors (LDRs)

Also called *resistive photodetectors*, these sensors operate with a modification of a physical property due to interaction with light. The surface of the sensing component of the sensor is made of a material which reacts to the light with a change in electron mobility and, as a consequence, in the material resistivity ρ. The choice of the material leads to different sensor characteristics, such as baseline value, sensitivity, activation threshold and so on.

Resistive gas sensors

Similarly to the devices already mentioned, resistive gas sensors show a the modification of material resistivity ρ caused by interaction with specific gases. The sensing component of these sensors is made of a material which reacts chemically with the gas molecules, increasing or decreasing the concentration of free electrons and, therefore, the surface resistivity. Semiconductors such as metal oxides are the sensing material usually employed for these sensors (thus, metal oxide sensors – MOX). The large variety of available materials guarantees a wide range of devices, each of which is characterized by selectivity towards one or more specific substances (Lundstrom, 1996; Meixner and Lampe, 1996). A MOX sensor is characterized by a steady-state resistance value, usually called the *baseline*, which depends on several parameters, such as the method or manufacture, material and working temperature. Baseline values of MOX sensors can therefore fall in a resistance interval of several decades. Moreover, the high sensitivity shown by MOX sensors to particular substances implies a modification of the sensor resistance value, which can also be in the order of several decades (Sberveglieri *et al.*, 2000). This makes the interface procedures quite critical, as will be clarified in subsequent sections.

Strain gauges

A resistive film is sustained by a thin and flexible support. The entire structure is tightly attached to an object and the deformation (tension and compression) of this object causes the deformation of the filament, in terms of length and cross-section. Thus, the geometrical portion of Equation [7.2] L/A is modified, leading to the modification of the filament resistance. If utilized in suitable structures, these sensors can be used to detect the strain of a material as well as strength, weight, pressure and so on.

Potentiometers

Potentiometers are the most straightforward resistive sensors. They are basically realized with a bar of length L of conductive material or wrapped conductive filament and a moving slide with a contact which touches the bar between its ends, as shown in Fig. 7.2. If the resistance of the whole bar (between the two ends) is R, the resistance R_x obtained between one terminal and the slide contact depends on the position x of the slide, with respect to the bar end, by means of the linear equation [7.3].

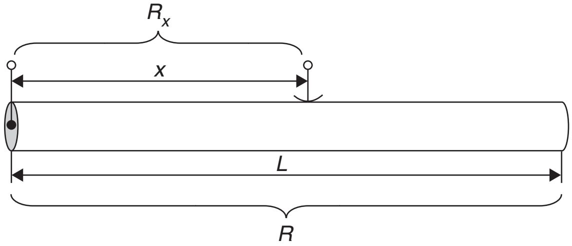
$$R_x = \frac{R}{L}x \qquad [7.3]$$

This kind of structure can be therefore used as a linear relative position sensor. It should be noted that, if a structure with a circular bar and a radial slide is implemented, an angular position sensor can be obtained.

Evident advantages are due to the simple structure and the linear relationship between the sensor input (x) and output (R_x); the main drawback is related to the sliding contact which, over time, 'consumes' the structure and determines a frequent need for recalibration or sensor replacement.

7.2.2 Parasitic capacitance

When dealing with the estimation of the resistance of resistive sensors, a specific problem to be resolved is related to the presence of capacitive effects.



7.2 Resistive linear position sensor based on potentiometer structure.

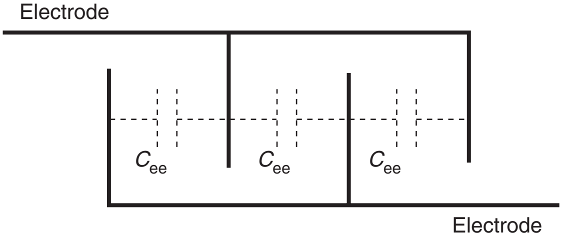
Unlike most of the circuit components' nonidealities, these capacitive effects cannot be compensated by a proper circuit calibration, because they depend on the particular sensor and operating conditions. If not appropriately taken into account, such effects can cause errors in the resistance estimation, as will be clarified in the next sections.

One of the most common origins of parasitic effects is related to sensor manufacture. If the sensing effect is obtained by certain phenomena occurring on the surface of the sensor (e.g., with photodetectors and gas sensors), the usual way to improve sensor sensitivity is to maximize the surface effects by implementing a technique based on interdigitated electrodes, shown in Fig. 7.3. Unfortunately, this structure introduces increased parasitic capacitance C_{ee} between the electrodes, as shown in Fig. 7.3, which becomes even more profound as the interdigitated structure is iterated (Olthuis *et al.*, 1997).

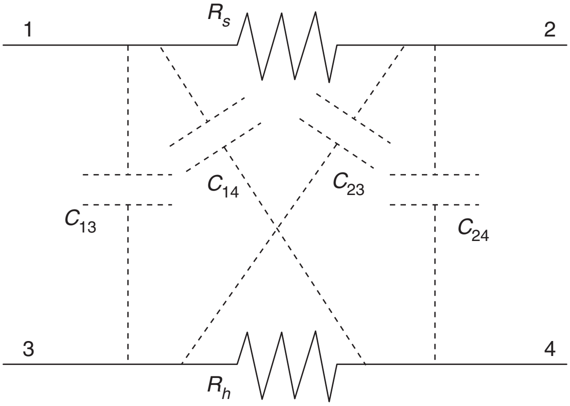
Another situation in which parasitic capacitive effects appear is particular to gas sensors. Some devices for gas sensing need to operate at a much higher temperature than the surrounding environment and, for this reason, they are usually provided with an embedded filament R_h which acts as the heater (Faglia *et al.*, 1999). The heater filament is a conductor realized on the same substrate of the sensing component R_s and separated by dielectric material that electrically isolates the two sensor components. However, the small size of the realized devices makes these two components interact with each other due to capacitive effects, as shown in Fig. 7.4.

In addition to these possible internal sources, capacitive effects can arise due to external reasons, such as the connection between the sensor and the measurement system, as shown in Fig. 7.5. In fact, the connectors and wires used to link the sensor to the electronic circuit show a distributed capacitive behaviour C_c which, from an instrumentation point of view, is seen in parallel with the sensor.

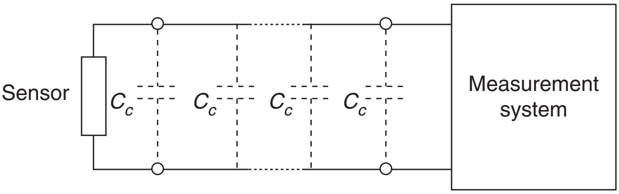
The computation of the overall capacitive parasitic effect is far from being easy and often involves a complete understanding of a sensor's characteristics, also at the microscopic level. For this reason, when



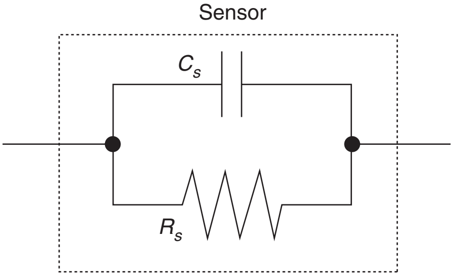
7.3 Parasitic capacitive effect in sensors with interdigitated electrodes.


1
R_s
2
C_{14}
C_{23}
C_{13}
C_{24}
3
4
R_h

7.4 Parasitic capacitive effect in gas sensors with embedded heater filament.


Sensor
C_c
C_c
C_c
C_c
Measurement system

7.5 Parasitic capacitive effect due to the sensor connections with the measurement system.


Sensor
C_s
R_s

7.6 Simplified model of a resistive sensor taking into account the parasitic capacitive effects.

parasitic capacitance needs to be taken into account, a simplified model of the sensor is usually considered, where a parasitic capacitor C_s is presented in parallel with the sensor resistance R_s, accounting for all possible capacitive parasitic effects. The simplified sensor model used in the next sections is shown in Fig. 7.6.

It should be pointed out that when R_s is very large (resembling the behaviour associated with an open circuit) C_s can predominate, thus leading to significant errors in the resistance estimation. In these cases (e.g., when dealing with MOX sensors), the interface circuits must be designed to limit this occurrence.

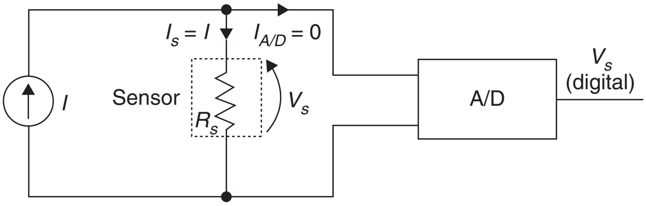
7.3 Voltamperometric resistance estimation

The most straightforward method with which to estimate the sensor resistance value is by means of the voltamperometric approach. Basically, a known current I is injected into the sensor, the voltage V_s across the sensor is measured by means of a voltmeter. Using Ohm's Law, the unknown sensor resistance R_s is calculated. The same technique can be used by applying a known voltage V and measuring the current I_s flowing through the sensor by means of an Amperemeter.

7.3.1 Implementation in smart sensors

When dealing with smart sensors, the measurement process is often performed using an A/D converter; this is either embedded in the microcontroller or a separate component. Since cost and simplicity of the electronic circuits and components are always a key factor, A/D converters with a voltage input are usually employed; thus, the first measurement modality, shown in Fig. 7.7, is advisable. If the input current $I_{A/D}$ of the A/D is considered to be zero, then the injected current I entirely flows through the sensor ($I_s = I$) and the sensor resistance can be estimated by:

$$R_s = \frac{V_s}{I_s} = \frac{V_s}{I} \qquad [7.4]$$



7.7 Resistance estimation with the voltamperometric method and a known current *I* injected to the sensor.

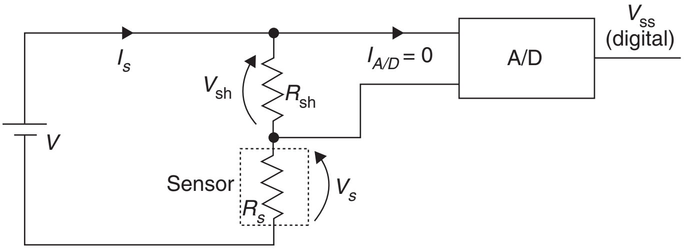
In this case, an excitation circuit able to inject the known current I into the sensor is needed. A simpler excitation circuit can be used to apply a known voltage V to the sensor; the second estimation method can be applied in this way, but only if a current/voltage conversion is performed – for instance, by means of a shunt resistor R_{sh}, as shown in Fig. 7.8. Once again, it is considered that $I_{\text{A/D}}$ is zero; thus, the generated current I_s flows both through the shunt resistor and the sensor, generating the V_{sh} and V_s voltages, respectively. The voltage V_{sh} is measured by means of the A/D converter and, since the R_{sh} is known, the sensor current I_s can be calculated using Ohm's Law. The sensor voltage V_s is obtained as $V_s = V - V_{\text{sh}}$ and, given V_s and I_s, the R_s value can be calculated using Ohm's Law. Equation [7.5] shows the final formula for the R_s computation:

$$R_s = \frac{V_s}{I_s} = \frac{R_{\text{sh}}}{V_{\text{sh}}}\left(V - V_{\text{sh}}\right) \qquad [7.5]$$

It should be noted that the voltage V_s applied to the sensor is not constant but, instead, depends on the sensor resistance value, according to Ohm's Law for the circuit in Fig. 7.7, and to the voltage divider relationship in Equation [7.6] for the circuit in Fig. 7.8.

$$V_s = V\frac{R_s}{R_s + R_{\text{sh}}} \qquad [7.6]$$

In most cases this is not a major issue, since the variation of R_s, and thus of V_s, has a limited range and the behaviour of the sensor is not significantly affected by the sensor polarization. However, especially when dealing with MOX sensors, variations in the sensor excitation voltage can cause the activation of inner phenomena; this leads to a sensor resistance drift



7.8 Resistance estimation with the voltamperometric method and a known voltage *V* applied to the sensor-shunt resistor series.

effect (Sauerwald *et al.*, 2005) and, thus, to unpredictable sensor behaviour and unreliable results.

An operational amplifier can be used in an architecture to overcome this problem, as shown in Fig. 7.9. Because of the virtual ground, the voltage V_s across the sensor is always equal to the applied and known voltage V independently of the R_s value. The output voltage V_o depends on the value of the feedback resistor R_f by means of Equation [7.7]; by measuring V_o and using Equation [7.7], the R_s value can thus be calculated.

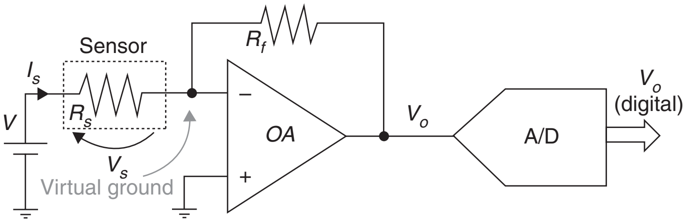
$$V_o = -V\frac{R_f}{R_s} \qquad [7.7]$$

The resolution of the A/D converter, which is related to the number of bits N, and the desired maximum relative error $\varepsilon_{\text{rel,max}}$ in the V_s measurement, determine the resistance range that can be estimated. Given the circuit in Fig. 7.7 and considering the full scale voltage V_{FS} of the A/D converter, the maximum resistance value $R_{\text{s,max}}$ which can be measured is given by Equation [7.8]; the minimum value $R_{\text{s,min}}$ is given by Equation [7.9].

$$R_{s,\max} = \frac{V_{FS}}{I} \qquad [7.8]$$

$$R_{s,\min} = \frac{V_{FS}}{2^N \varepsilon_{\text{rel,max}} I} = \frac{R_{s,\max}}{2^N \varepsilon_{\text{rel,max}} I} \qquad [7.9]$$

If a 1% resolution in the measurement of V_s is required (i.e., $\varepsilon_{\text{rel,max}} = 1\%$) and a 16-bit A/D converter is used, the operating range (i.e., the ratio between $R_{\text{s,max}}$ and $R_{\text{s,min}}$) of the circuit in Fig. 7.7 is less than three decades. To increase the resistance range to four decades and retain the same measurement

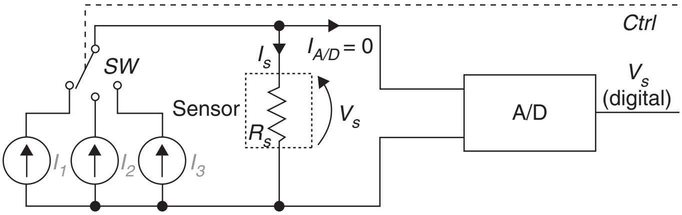


7.9 Resistance estimation with the voltamperometric method and a known voltage *V* applied to the sensor.

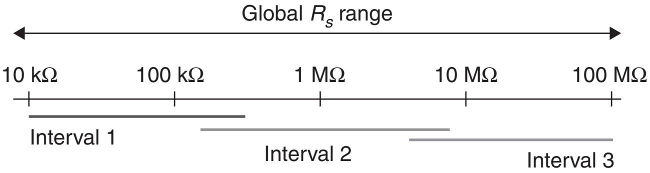
performance, a 20-bit A/D converter is needed, significantly increasing the system cost. To increase the range further to six decades, a 27-bit A/D converter would be required, thus making this approach impracticable.

The typical solution to enlarge the resistance range is to adopt a multiple interval technique. Referring to the circuit in Fig. 7.7, if the excitation current *I* is varied, the R_s range is shifted up or down. Thus, a set of appropriate *I* values is chosen in order to create multiple R_s intervals; in each of them, the A/D converter provides the measurement with the desired resolution. The multiple intervals are designed to cover, possibly with some overlaps, the complete desired R_s range, which can be significantly wider than the single interval. A circuit adopting this approach is presented in Fig. 7.10, whereas Fig. 7.11 presents an example of coverage of a four-decade R_s range by using the circuit in Fig. 7.10 with the intervals designed to cover about one-and-a-half decades each. The multiple interval technique can also be adopted with the circuit in Fig. 7.9, by using multiple feedback resistors, as shown in Fig. 7.12 (Grassi *et al.*, 2005).

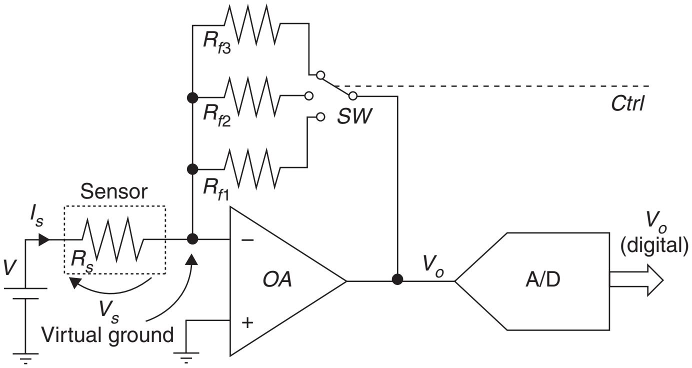
In both cases, the switch *SW* selecting the active sensor current value or feedback resistor is driven by a control signal *Ctrl* which has to be provided by the microcontroller or programmable logic device (PLD) of the smart sensor. This control must be performed by tracking the current value of the R_s and deciding which interval is the most suitable to acquire the resistive value.



7.10 The circuit in Fig. 7.7 modified to expand the measurable resistance range by using the multiple interval approach.



7.11 Example of the application of the circuit in Fig. 7.10 covering a four-decade resistance range by using three intervals each with a width of one-and-a-half decades.



7.12 The circuit in Fig. 7.9 modified to expand the measurable resistance range by using the multiple interval approach.

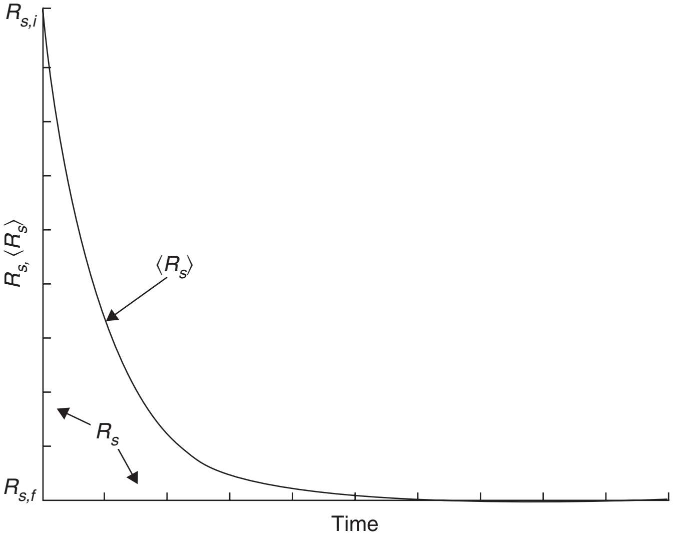
The interval overlap regions are important for the calibration process. In fact, it must be guaranteed that, if the same resistance value can be measured using two or more intervals, the results are equal within the measurement uncertainty. If this cannot be guaranteed, nonlinear effects such as hysteresis and step variations of the estimated R_s can occur during the interval commutation due to resistance variation. The calibration procedure for this kind of circuit is therefore critical and may require significant resources, both in terms of time and costs. The adoption of this measurement technique for low-cost smart sensors therefore needs to be evaluated carefully.

7.3.2 Parasitic capacitance issues

Parasitic capacitive effects, due to the connection cables or the sensor itself, can in some cases contribute to the R_s estimation being unreliable. Some considerations for the use of the solutions previously discussed will be now be offered for a sensor model with resistance R_s in parallel with capacitance C_s.

First, it should be noted that if the sensor voltage V_s has been constant for some considerable time before the R_s measurement, as is the case with the circuit in Fig. 7.9, the parasitic capacitance C_s is not recharged and therefore it does not affect the measurement. Conversely, when using the circuit in Figs 7.7 and 7.8, V_s varies with the R_s variation, thus the contribution of C_s to the measurement should be taken into account. In the following, some examples will be given to show the typical measurement problems encountered because of C_s.

Defining $\langle R_s \rangle$ as the estimated value of R_s obtained with the circuit in Fig. 7.7 and Equation [7.4], a simulation of the behaviour of $\langle R_s \rangle$ vs time is reported in Fig. 7.13, when a step variation of R_s (with final value $R_{s,f}$ lower than the initial value $R_{s,i}$) is analysed. The measured $\langle R_s \rangle$ follows a



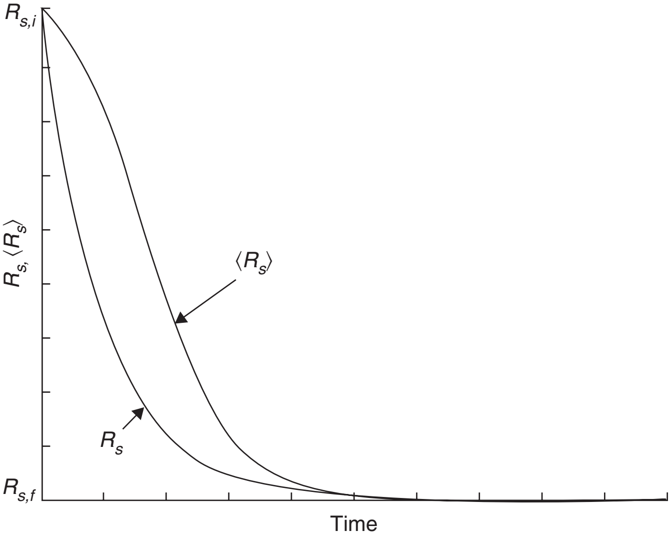
7.13 Simulation of the circuit in Fig. 7.7 when the sensor resistance has a step variation and the parasitic capacitive effects are taken into account.

descending exponential line starting from $R_{s,i}$, towards $R_{s,f}$ and with time constant $\tau = R_{s,f} \cdot C_s$. The final effect is that the measurement is a slowed down version of the real R_s behaviour, and this effect is progressively more visible as the parasitic capacitive effect grows. In addition, it should be noted that this effect also becomes more evident with an increase in the R_s value, thus circuits operating with a wide resistance range do not offer a uniform performance.

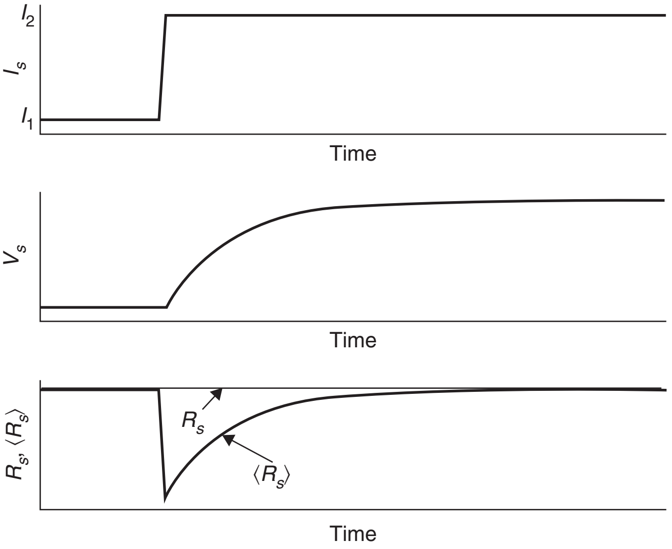
Figure 7.14 shows a simulation of the circuit in Fig. 7.8, where the sensor resistance has a descending exponential behaviour, typical of MOX gas sensors. Also in these circumstances, the effect of C_s implies that the estimated resistance value $\langle R_s \rangle$ appears slower than the real R_s.

It should be pointed out that, for applications in which the variation of the sensor resistance is slower than the delays introduced by the parasitic capacitance, this effect could be ignored. Conversely, if the information to be extracted from the sensor is related not only to the R_s value, but also to the dynamics of the R_s variation, the presence of parasitic capacitance effects could significantly alter the measurement results and must therefore be taken into account.

Figure 7.15 shows a simulation of the circuit in Fig. 7.10, where a steady-state resistance R_s is considered and the injected current I_s varies, because of a scale change, from I_1 to I_2, with $I_2 > I_1$. Since R_s is constant and the I_s current increases in a step variation, according to Ohm's Law the V_s voltage should also increase in the same quantity and with the same



7.14 Simulation of the circuit in Fig. 7.8 when the sensor resistance has a descending exponential variation and the parasitic capacitive effects are taken into account.



7.15 Simulation of the circuit in Fig. 7.10 when the sensor resistance is constant, a scale exchange is applied and the parasitic capacitive effects are taken into account.

behaviour. However, the presence of C_s makes the sensor voltage V_s follow an exponential law with time constant $\tau = R_s \cdot C_s$, when reaching its new value. The effect on the estimated sensor resistance $\langle R_s \rangle$, evaluated by means of Equation [7.4], is a glitch that corresponds with the scale change, as visible in the third graph of Fig. 7.15, the width of which is proportional to R_s and C_s. Once again, according to the specific application and to the involved R_s and C_s values, such an effect can be either insignificant or unacceptable.

Finally, a brief analysis of the measurements using a sinusoidal sensor excitation voltage/current, instead of constant values, will be given. Using a parallel presentation $R_s//C_s$ of the sensor impedance Z_s, Equation [7.10] shows the expression of Z_s in terms of real and imaginary part, whereas Equation [7.11] shows the expression of Z_s in terms of magnitude and phase, where ω is the angular speed of the excitation signal.

$$Z_s = \frac{R_s}{1+(\omega R_s C_s)^2} - j\frac{\omega R_s^2 C_s}{1+(\omega R_s C_s)^2} \qquad [7.10]$$

$$|Z_s| = \frac{R_s}{\sqrt{1+(\omega R_s C_s)^2}}; \quad \angle(Z_s) = \arctan(-\omega R_s C_s) \qquad [7.11]$$

The use of an impedance analyser allows the real and imaginary part (or, similarly, the magnitude and phase) of Z_s to be estimated and, therefore, both the R_s and C_s value to be calculated. In smart sensor implementations, this implies that the A/D acquisition needs to be synchronized with the sensor excitation voltage/current in order to evaluate the Z_s phase correctly by means of the input/output signals phase shift. In the event that only the Z_s magnitude is evaluated, an underestimation of R_s is obtained, since it is evident from Equation [7.11] that $|Z_s| < R_s$ for each ω value and that this error becomes progressively greater as the values of ω and C_s increase.

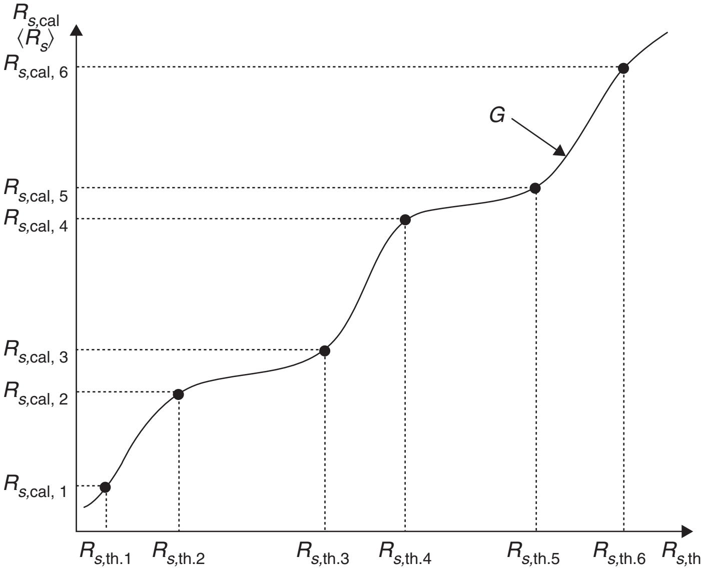
7.3.3 Calibration procedures

The calibration of a system involving resistive sensors does not differ from the usual procedures. When dealing with smart sensors, the presence of a processing unit such as the microcontroller or PLD allows these techniques to be implemented directly on the sensor node, thus facilitating the communication, sharing and utilization of the sensor data. Moreover, additional circuitry and automatic routines could be implemented to perform online and periodic recalibration of the system in order to compensate for long-term effects such as component aging (Grassi *et al.*, 2005).

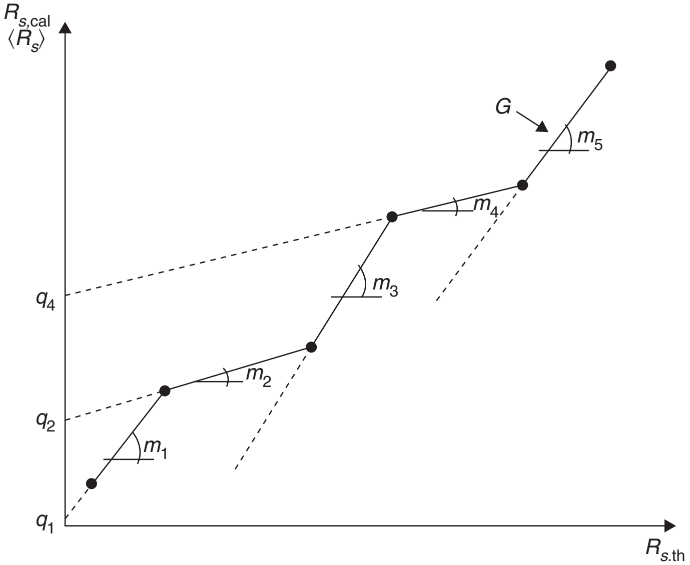
Basically, given a set of known R_s values $R_{s,\mathrm{cal}}$, the corresponding calculated resistance values $R_{s,\mathrm{th}}$ (obtained by means of the theoretical circuit equations starting from the measured quantities – e.g., voltage V_s in Fig. 7.7) usually differ from $R_{s,\mathrm{cal}}$, because of circuit imperfections and component nonidealities. An experimentally designed calibration function G should be applied to the theoretical resistance values $R_{s,\mathrm{th}}$ to obtain, as the final result $\langle R_s \rangle$ of the estimation, the calibration values $R_{s,\mathrm{cal}}$, thus compensating for the circuit errors. An example of a sophisticated calibration function G is shown in Fig. 7.16. It is no problem to implement such a function by means of a cheap microcontroller or look-up table stored in a PLD. However, collecting a sufficient amount of calibration data to design an efficient calibration function is still a long and expensive procedure.

If calibration procedures need to be simplified, a linearization of the calibration function is advisable. Basically, with a very small number of calibration points, the calibration function can be approximated by a single line or a set of lines.

Figure 7.17 shows an example of a piecewise linearization applied to a set of N calibration samples $R_{s,\mathrm{cal}}$. G is realized with multiple lines ($N - 1$), each of which characterized by the slope m and offset q, connecting each pair of consecutive calibration samples. The m and q parameters of each



7.16 Example of a complete calibration function G obtained with a set of calibration samples $R_{s,\mathrm{cal}}$.



7.17 Example of a calibration function *G* obtained with a piecewise linearization starting from a set of calibration samples $R_{s,\mathrm{cal}}$.

segment *i* (*i* = 1, …, *N*) can be easily defined from the calibration samples $R_{s,\mathrm{cal}}$ and the related calculated values $R_{s,\mathrm{th}}$ with Equations [7.12] and [7.13], respectively.

$$m_i = \frac{R_{s,\mathrm{cal},i} - R_{s,\mathrm{cal},i-1}}{R_{s,\mathrm{th},i} - R_{s,\mathrm{th},i-1}} \qquad [7.12]$$

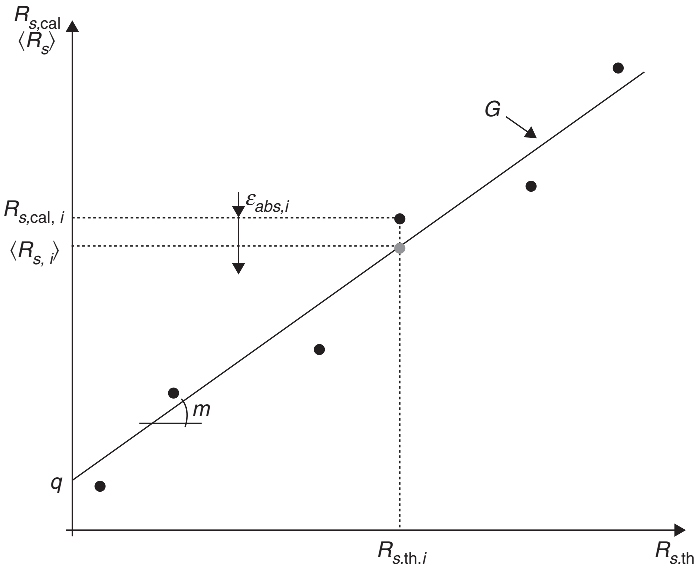
$$q_i = R_{s,\mathrm{cal},i} - m_i R_{s,\mathrm{th},i} \qquad [7.13]$$

To obtain the final result $\langle R_s \rangle$ of the measurement, the microcontroller or the PLD of the smart sensor needs to store the pairs of slope *m* and offset *q* of each line in a look-up table and apply the linear relation in Equation [7.14] with the calculated values $R_{s,\mathrm{th}}$, selecting the correct *m*-*q* pair according to the $R_{s,\mathrm{th}}$ itself.

$$\langle R_s \rangle = mR_{s,\mathrm{th}} + q \qquad [7.14]$$

A single line can also be used to further simplify the calibration function.

Figure 7.18 shows an example of approximation using a linear regression or least mean squares (LMS) algorithm. The single line approximating the



7.18 Example of a calibration function *G* obtained with a linear regression starting from a set of calibration samples $R_{s,\mathrm{cal}}$.

calibration function is defined to minimize the sum of the absolute errors ($\varepsilon_{\mathrm{abs}} = \langle R_s \rangle - R_{s,\mathrm{cal}}$) of the estimated values $\langle R_s \rangle$ with respect to the related $R_{s,\mathrm{cal}}$. Given N calibration samples $R_{s,\mathrm{cal}}$ and the related calculated values $R_{s,\mathrm{th}}$, the m and q values can be obtained by Equations [7.15] and [7.16], respectively.

$$m = \frac{N\sum_i \left(R_{s,\mathrm{th},i} R_{s,\mathrm{cal},i}\right) - \left(\sum_i R_{s,\mathrm{th},i}\right)\left(\sum_i R_{s,\mathrm{cal},i}\right)}{N\sum_i \left(R_{s,\mathrm{th},i}\right)^2 - \left(\sum_i R_{s,\mathrm{th},i}\right)^2} \qquad [7.15]$$

$$q = \frac{\left(\sum_i R_{s,\mathrm{cal},i}\right)\sum_i \left(R_{s,\mathrm{th},i}\right)^2 - \left(\sum_i R_{s,\mathrm{th},i}\right)\left(\sum_i \left(R_{s,\mathrm{th},i} R_{s,\mathrm{cal},i}\right)\right)}{N\sum_i \left(R_{s,\mathrm{th},i}\right)^2 - \left(\sum_i R_{s,\mathrm{th},i}\right)^2} \qquad [7.16]$$

From a smart sensor perspective, the microcontroller or the PLD simply needs to apply the linear equation [7.14] with the calculated values $R_{s,\mathrm{th}}$ to obtain the final result $\langle R_s \rangle$ of the estimation.

This approach is particularly advantageous when the sensor resistance has a limited variation range; for instance, two or three decades. When the variation range becomes too great – for example, when dealing with MOX sensors – a more important parameter to be considered for the linearization of the calibration function is the relative error, rather than the absolute error. In fact, minimizing the absolute error in a wide range of resistance would penalize the samples with smaller values, leading to a considerable error compared with the resistance value itself. The relative error is defined as the absolute error of the estimation referred to the 'true value' expected from the measurement; that is, $\varepsilon_{\mathrm{rel}} = \varepsilon_{\mathrm{abs}}/R_{s,\mathrm{cal}}$. If the m and q of the calibration line are estimated to minimize this parameter, the overall performance of the calibration procedure, in terms of relative error, is more uniform in the range under consideration.

This linearization technique is called weighted LMS (WLMS); Equations [7.17] and [7.18] can be used to obtain the values of the parameters m and q:

$$m = \frac{\left(\Sigma_i \frac{R_{s,\mathrm{th},i}}{R_{s,\mathrm{cal},i}}\right)\left(\Sigma_i \frac{1}{\left(R_{s,\mathrm{cal},i}\right)^2}\right) - \left(\Sigma_i \frac{1}{R_{s,\mathrm{cal},i}}\right)\left(\Sigma_i \frac{R_{s,\mathrm{th},i}}{\left(R_{s,\mathrm{cal},i}\right)^2}\right)}{\left(\Sigma_i \frac{1}{\left(R_{s,\mathrm{cal},i}\right)^2}\right)\left(\Sigma_i \left(\frac{R_{s,\mathrm{th},i}}{R_{s,\mathrm{cal},i}}\right)^2\right) - \left(\Sigma_i \frac{R_{s,\mathrm{th},i}}{\left(R_{s,\mathrm{cal},i}\right)^2}\right)} \qquad [7.17]$$

$$q = \frac{\left(\Sigma_i \frac{1}{R_{s,\mathrm{cal},i}}\right)\left(\Sigma_i \left(\frac{R_{s,\mathrm{th},i}}{R_{s,\mathrm{cal},i}}\right)^2\right) - \left(\Sigma_i \frac{R_{s,\mathrm{th},i}}{R_{s,\mathrm{cal},i}}\right)\left(\Sigma_i \frac{R_{s,\mathrm{th},i}}{\left(R_{s,\mathrm{cal},i}\right)^2}\right)}{\left(\Sigma_i \frac{1}{\left(R_{s,\mathrm{cal},i}\right)^2}\right)\left(\Sigma_i \left(\frac{R_{s,\mathrm{th},i}}{R_{s,\mathrm{cal},i}}\right)^2\right) - \left(\Sigma_i \frac{R_{s,\mathrm{th},i}}{\left(R_{s,\mathrm{cal},i}\right)^2}\right)^2} \qquad [7.18]$$

Table 7.1 shows an example of calibration on a five-decade resistive range, comparing the absolute and relative errors in case of no calibration, LMS and WLMS linearization. Where there is no calibration, the $R_{s,\mathrm{th}}$ is the final result of the measurement, whereas in the other two cases, the final result of the estimation is $\langle R_s \rangle$, computed by means of Equation [7.14]. It should be noted how the LMS calibration tries to make uniform the absolute error in the whole range under consideration, leading to significantly high relative error values in the lower part of the range. Conversely, the WLMS approach makes the relative error as uniform as possible, accepting significantly high absolute error values in the upper part of the range.

Table 7.1 Comparative table of the calibration results on a five-decade resistance range where there is no calibration, LMS and WLMS linearization

$R_{s,\mathrm{cal}}$ (MΩ)	$R_{s,\mathrm{th}}$ (MΩ)	No calibration		LMS			WLMS		
		$\varepsilon_{\mathrm{abs}}$ (MΩ)	$\varepsilon_{\mathrm{rel}}$ (%)	$\langle R_s \rangle$ (MΩ)	$\varepsilon_{\mathrm{abs}}$ (MΩ)	$\varepsilon_{\mathrm{rel}}$ (%)	$\langle R_s \rangle$ (MΩ)	$\varepsilon_{\mathrm{abs}}$ (MΩ)	$\varepsilon_{\mathrm{rel}}$ (%)
1	0.90	0.10	10.00	177.93	–176.93	–17693.43	0.99	0.01	0.52
10	11.00	–1.00	–10.00	187.15	–177.15	–1771.48	10.61	–0.61	–6.15
100	95.00	5.00	5.00	263.77	–163.77	–163.77	90.62	9.38	9.38
1000	1080.00	–80.00	–8.00	1162.32	–162.32	–16.23	1028.82	–28.82	–2.88
10 000	9950.00	50.00	0.50	9253.78	746.22	7.46	9477.31	522.69	5.23
100 000	109 500.00	–9500.00	–9.50	100 066.05	–66.05	–0.07	104 296.68	–4296.68	–4.30

7.4 Resistance-to-time conversion methods

The measurement techniques illustrated in this section are particularly advantageous when the resistance value needs to be estimated over a very wide range. This can be the case for a single sensor (the value of which can span a large interval), as well as of a set of sensors (each of which having a limited resistance variation, but centred around different reference values). A typical example of such a situation is represented by MOX sensors.

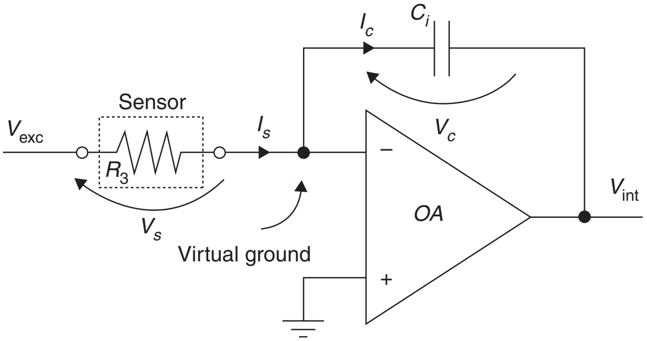
To assure simple and inexpensive calibration of the electronic circuits for the sensor interface, the schemes adopting multiple ranges for the estimation of resistance should be avoided. However, as previously stated, implementing a direct resistance measurement operating on a wide range and guaranteeing the desired performance over the whole interval can be either cost-inefficient or impracticable.

The resistance-to-time conversion (RTC) method is based on the fact that the quantity 'time' is easy to estimate, also in a wide range, with good resolution and by using relatively inexpensive components, such as microcontrollers and PLDs.

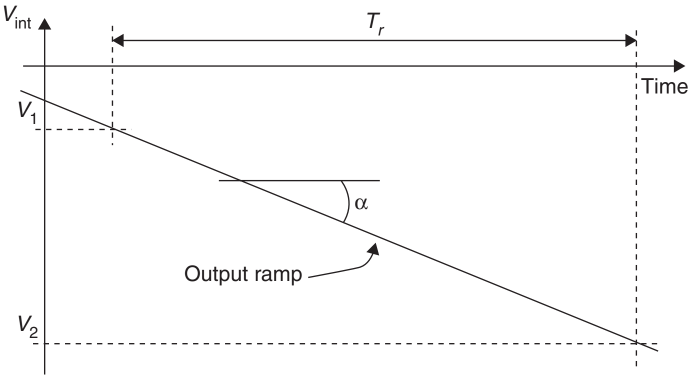
The basic element used for RTC is an integrator, shown in Fig. 7.19. Considering *OA* as an ideal operational amplifier, it can easily be found that $I_s = I_c$, $V_s = V_{\text{exc}}$ and $V_{\text{int}} = -V_c$, thus yielding Equation [7.19] describing the behaviour of the circuit.

$$V_{\text{int}}(t) = -V_c(t) = -\frac{1}{C_i}\int I_c(t)\,dt = -\frac{1}{R_sC_i}\int V_{\text{exc}}(t)\,dt \qquad [7.19]$$

Under the hypothesis that the sensor excitation voltage V_{exc} and the sensor resistance R_s are constant, the integrator output voltage V_{int} is a ramp, as illustrated in Fig. 7.20, the slope α of which is inversely proportional to the sensor resistance value, as shown by Equation [7.20].



7.19 The integrator circuit, the basic element for the resistance-to-time conversion.



7.20 Output time diagram of the integrator in Fig. 7.19, considering an ideal operational amplifier and constant values for V_{exc} and R_s.

$$V_{int}(t) = -\frac{V_{exc}}{R_s C_i} t + V_i = -\alpha t + V_i \quad \text{with} \quad \alpha = \frac{V_{exc}}{R_s C_i} \quad \text{and} \quad V_i = V_0 \ (t = 0) \qquad [7.20]$$

The time T_r taken by the output ramp to span between two known voltage values V_1 and V_2 is inversely proportional to the slope α of V_{int}, thus a direct proportional relationship can be found between T_r and the sensor resistance R_s, as in Equation [7.21].

$$\alpha = \frac{V_{exc}}{R_s C_i} = \frac{|V_2 - V_1|}{T_r} \qquad [7.21]$$

The measure of the time T_r is therefore the key point for the estimation of sensor resistance. To accomplish this task, the integrator circuit is utilized in different schemes, which mainly differ for the methodology adopted to extract the information T_r, to iterate the measurement and for the way the sensor is biased (i.e., constant or switched voltage). Particularly, in this latter case the problems related to the parasitic capacitance are more evident (Sauerwald *et al.*, 2005). As it will be shown next, this issue can be limited by suitably designing the electronic interface.

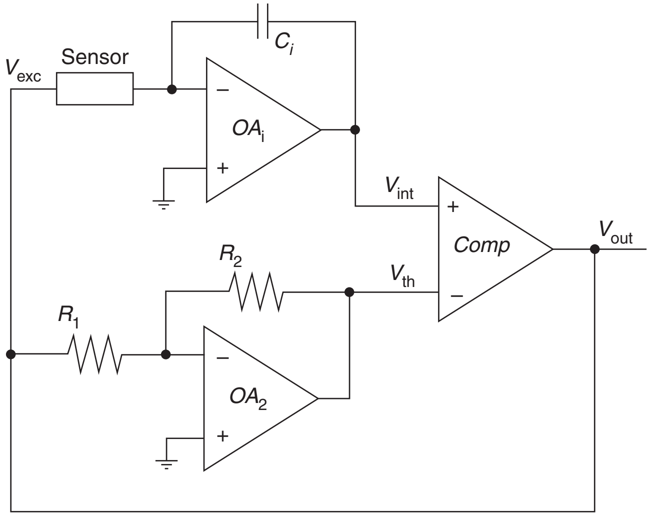
Usually, the analog front-end provides a quasi-digital signal, containing the information T_r. The measurement of T_r is accomplished with a digital system, which can be a digital counter or, in case of smart sensors, time measurement routines or blocks implemented in a microcontroller or PLD.

In the following, an overview of different RTC-based interface circuits will be presented, highlighting the main advantages and drawbacks for each of them. Unless otherwise specified, the sensor model including the parasitic capacitive effect C_s in parallel with the resistive value R_s will be used.

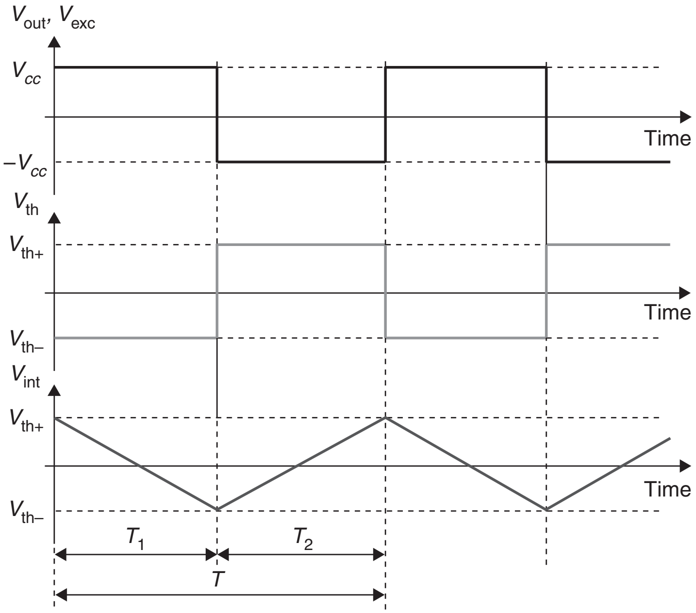
It should be noted that all the resistance-to-time methods are based on the use of a capacitor for the sensor current integration; since the capacitive value of common capacitors is usually not very accurate and stable, the accuracy of the measurement system can be affected as well. As will become clear in the following, the capacitive value is usually a multiplicative factor in the relationship for the sensor resistance calculation, and therefore the initial calibration/linearization procedure can significantly limit the problem. Moreover, it should be remembered that the proposed sensor interfaces are intended to be used in smart sensor systems, where a microcontroller or PLD could easily implement periodic and simple self-recalibration procedures to correct possible drifts of the circuit component values, included the integrating capacitor.

7.4.1 Oscillator-based systems

This family of circuits employs a first-order oscillator scheme to charge and discharge the capacitor of the integrator continuously, thus offering an easy way to iterate the measurement (Flammini *et al.*, 2004). The basic circuit and its timing diagram are shown in Figs 7.21 and 7.22, respectively. In this first analysis, the sensor is considered as a pure resistor R_s, thus neglecting possible parasitic capacitive effects. The sensor is biased with the output voltage of the comparator *Comp* ($V_{exc} = V_{out}$), thus it is an alternate voltage commutating between the power supply value $\pm V_{cc}$ (without loss of generality, it is supposed that rail-to-rail components are used). The capacitor C_i of the integrator is consecutively recharged, thus generating the triangle waveform



7.21 Basic scheme of the oscillating circuit for the resistance-to-time conversion.



7.22 Timing diagram of the circuit in Fig. 7.21.

at the integrator output V_{int}, as illustrated in Fig. 7.22. The inverting amplifier, composed of the operational amplifier OA_2, R_1 and R_2, provides two alternative threshold values V_{th+} and V_{th-} by inverting and amplifying the comparator output V_{out}. In this way, as can be seen in Fig. 7.22, first-order oscillator behaviour is established.

The duration of the time intervals T_1 and T_2 depends on the circuit parameters and on the value of the sensor resistance R_s, as in Equation [7.22].

$$T_1 = T_2 = C_i R_s \frac{|V_{th+} - V_{th-}|}{V_{exc}} = 2GC_i R_s \quad \text{with} \quad G = \frac{R_2}{R_1} \qquad [7.22]$$

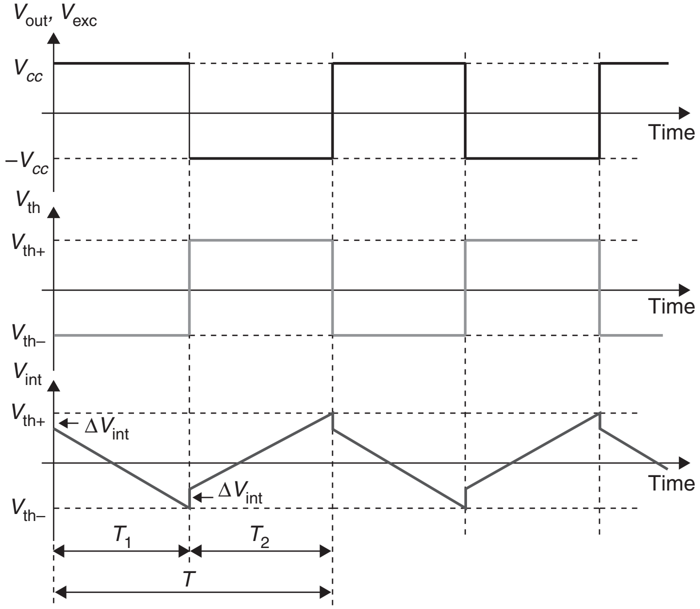
Using this method, for every period of V_{out} it is possible to derive two values of R_s by measuring the time intervals T_1 and T_2. It is worth noting that Equation [7.22] has been obtained by considering an ideal case, with ideal components, symmetric power supply voltage and homogeneous sensor behaviour, with respect to the alternate excitation voltage V_{exc}. Partial compensation for any unwanted effects can be achieved by averaging the two half periods, by measuring the period T and calculating the sensor resistance R_s using Equation [7.23].

$$R_s = \frac{T}{4GC_i} \quad \text{with} \quad G = \frac{R_2}{R_1} \qquad [7.23]$$

However, it should be noted that the integrator output waveform and Equations [7.22] and [7.23] have been obtained supposing that the resistance R_s is constant (or slowly varying) during the observation time. Thus, the design parameters of the circuit, on which the observation time depends, should take into account the dynamic behaviour of the sensor resistance. Basically, to consider R_s constant during the measurement, the maximum variation of the sensor resistance during the observation time should be within the measurement uncertainty. If the hypothesis of constant R_s cannot be achieved, Equation [7.23] provides an estimation of an average value of R_s within the observation time T.

Parasitic capacitance issues

Since the sensor excitation voltage V_{exc} is an alternate voltage, the presence of parasitic capacitance in parallel to R_s could lead to significant errors. As illustrated in Fig. 7.23, during the commutation of V_{exc}, the integrator output V_{int} has an instantaneous variation ΔV_{int}, due to a charge transfer effect



7.23 Timing diagram of the circuit in Fig. 7.21 with non-negligible sensor parasitic capacitive effects.

involving the capacitors C_i and C_s. This voltage step is proportional to the parasitic capacitance C_s, as in Equation [7.24].

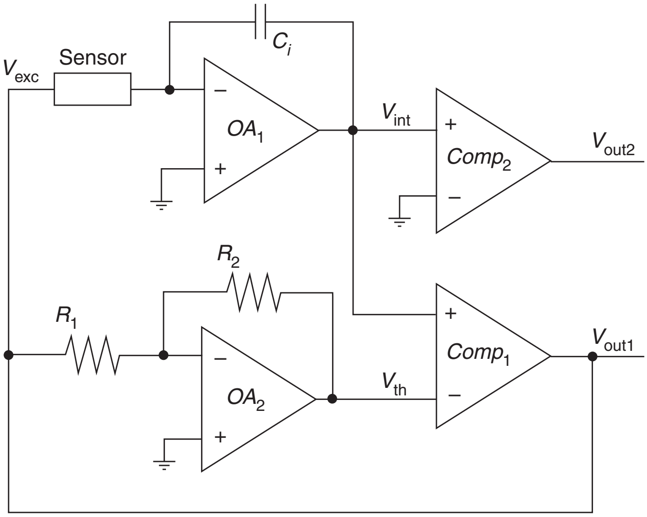
$$|\Delta V_{\text{int}}| = |\Delta V_{\text{exc}}| \frac{C_s}{C_i} \tag{7.24}$$

The main effect on the circuit output signal V_{out} is the shortening of the time intervals T_1 and T_2 (and therefore of the period T), with respect to the case in which the parasitic capacitance is neglected, as reported in Equation [7.25].

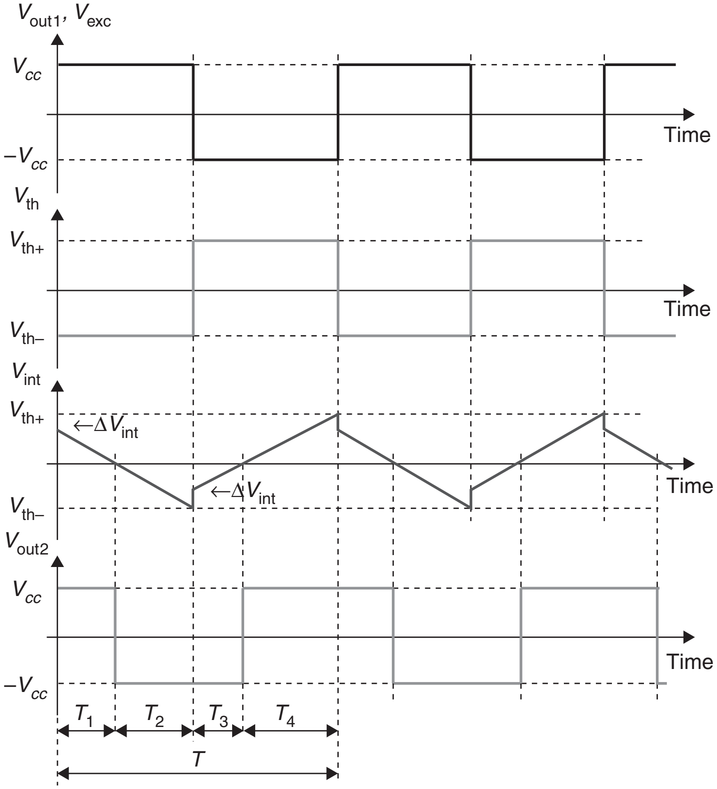
$$T_1 = T_2 = 2\left(G - \frac{C_s}{C_i}\right)C_i R_s \quad \text{with} \quad G = \frac{R_2}{R_1} \tag{7.25}$$

If T_1 and T_2 (or T) are measured and Equation [7.23] is applied, the sensor resistance R_s is obtained with an underestimation error, which becomes progressively more significant as the parasitic effect C_s becomes comparable with the integration capacitor C_i.

To overcome this issue, the circuit in Fig. 7.21 can be modified by adding a second comparator, operating with a different threshold voltage at its input, as shown in Fig. 7.24 (Ferri *et al.*, 2008). In this case, the second threshold value has been chosen to be in the middle of the integrator output range. This is the ground voltage in the event of a symmetrical power supply



7.24 The modified oscillating circuit for the resistance-to-time conversion for non-negligible sensor parasitic capacitive effects.



7.25 Timing diagram of the circuit in Fig. 7.24.

$\pm V_{cc}$. The timing diagram of the circuit in Fig. 7.24 is reported in Fig. 7.25. A new square wave output signal (V_{out2}) is provided; the commutations of V_{out2} occur when the integrator output crosses the second threshold value (ground voltage). During a period T of the signals, four time intervals (T_1, T_2, T_3 and T_4) can be defined, each of which is delimited by the commutations of V_{out1} and V_{out2}.

As can be seen in Fig. 7.25, the time intervals T_2 and T_4 are not affected by the presence of the parasitic capacitive effects, being dependent only on the threshold voltage values and on the slope of V_{int}. Thus, the measurement of such time intervals allows the sensor resistance to be estimated without being influenced by the parasitic capacitive effects. Equation [7.26] shows the relationship between T_2 (T_4) and R_s in an ideal situation, whereas Equation [7.27] reports how R_s can be computed by averaging the information obtained from T_2 and T_4, thus partially compensating possible circuit asymmetries.

$$T_2 = T_4 = 2GC_iR_s \quad \text{with} \quad G = \frac{R_2}{R_1} \qquad [7.26]$$

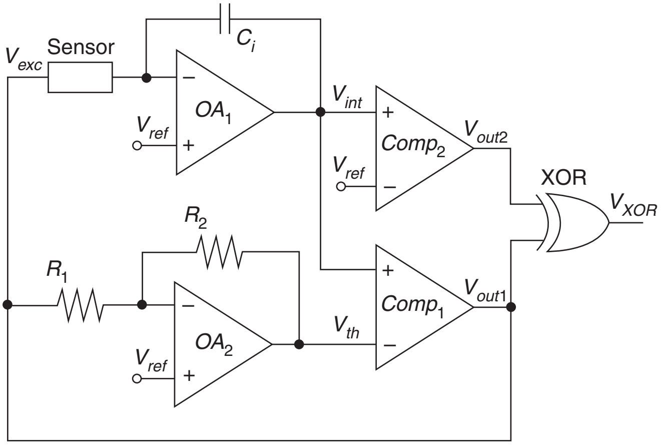
$$R_s = \frac{T_2 + T_1}{2GC_t} \quad \text{with} \quad G = \frac{R_2}{R_1} \qquad [7.27]$$

It should be noted that, from the additional measurement of the time intervals T_1 and T_3 and by applying the relationship in Equation [7.24], an estimation of the parasitic capacitance C_s can be obtained, as in Equation [7.28].

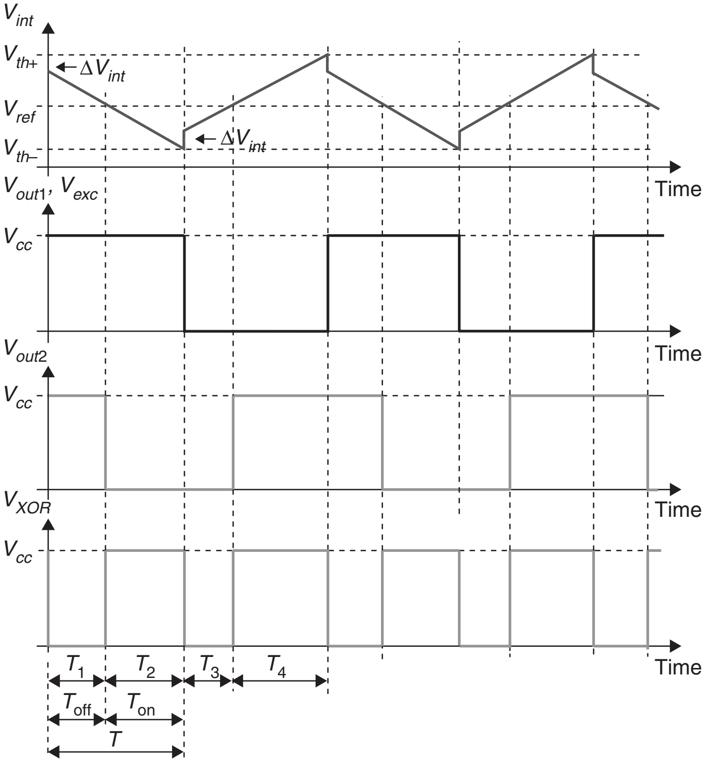
$$C_s = GC_i \frac{T_2 + T_4 - T_1 - T_3}{2(T_2 + T_4)} \quad \text{with} \quad G = \frac{R_2}{R_1} \qquad [7.28]$$

The estimation of C_s could be used for diagnostic purposes, for example to monitor the effectiveness of long connections between the sensor and the electronic front-end. This task can be easily performed by the microcontroller or the PLD devoted to the smart sensor management, thus including the measurement of the time intervals T_1, T_2, T_3 and T_4. In particular, for the latter task, the digital device needs to acquire the two logical outputs V_{out1} and V_{out2}. The implementation of counter-routines in microcontrollers requires dedicated ports with input capture features. However, if it is intended to use a cheap microcontroller, it is possible that this is available for one input only. In this case, the signal obtained from the Exclusive OR (XOR) logic combination of V_{out1} and V_{out2} can be used as a single output of the analog front-end, as shown next. Moreover, to facilitate the integration of the analog front-end with the digital stage for a possible single-chip solution, the circuit in Fig. 7.24 can be designed to operate with a single-supply voltage V_{cc}. This is obtained by replacing all the connections to ground voltage with a reference voltage V_{ref} placed in the middle of the V_{int} range; that is, $V_{\text{ref}} = V_{\text{cc}}/2$ (De Marcellis *et al.*, 2008). Figures 7.26 and 7.27 show the electronic interface circuit with the two modifications in place and the related timing diagram.

In an ideal situation, $T_1 = T_3$, $T_2 = T_4$ and the circuit output V_{XOR} is a square wave, the period of which is given by $T = T_1 + T_2 = T_3 + T_4$. Defining T_{on}, the time interval during which V_{XOR} is at the logic level 'high', and T_{off}, the time interval during which V_{XOR} is at the logic level 'low', it is evident that $T_{\text{on}} = T_2 = T_4$ and $T_{\text{off}} = T_1 = T_3$. The duty-cycle of V_{XOR} is then defined as $D = T_{\text{on}}/T$. From the estimation of T and D of the single output V_{XOR}, it is


V_{exc}
Sensor
C_i
OA_1
V_{ref}
V_{int}
$Comp_2$
V_{out2}
V_{ref}
XOR
V_{XOR}
R_2
R_1
$Comp_1$
V_{out1}
OA_2
V_{th}
V_{ref}

7.26 The oscillating circuit with single power supply and single output signal.


V_{int}
V_{th+}
ΔV_{int}
V_{ref}
V_{th-}
ΔV_{int}
Time
V_{out1}, V_{exc}
V_{cc}
Time
V_{out2}
V_{cc}
Time
V_{XOR}
V_{cc}
Time
T_1
T_2
T_3
T_4
T_{off}
T_{on}
T

7.27 Timing diagram of the circuit in Fig. 7.26.

possible to estimate the values of the sensor resistance R_s and the parasitic capacitance C_s by means of Equations [7.29] and [7.30], respectively.

$$R_s = \frac{TD}{GC_t} \quad \text{with} \quad G = \frac{R_2}{R_1} \tag{7.29}$$

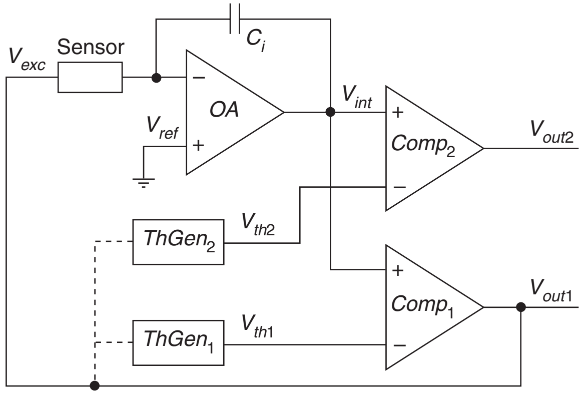
$$C_s = \frac{GC_t}{2}\left(2 - \frac{1}{D}\right) \quad \text{with} \quad G = \frac{R_2}{R_1} \tag{7.30}$$

It should be noted that the behaviour of V_{out1} and V_{out2} is identical as that in the previous case; therefore, Equations [7.27] and [7.28] can still be used, if the digital stage is able simultaneously to acquire the two output signals V_{out1} and V_{out2} and to measure the time intervals T_1, T_2, T_3 and T_4 appropriately. However, the use of the single output V_{XOR} can simplify the implementation of the digital counter-routines in the microcontroller, since a simple period and duty-cycle measurement needs to be performed. Such measurement tools are often included as a hardware feature in most microcontrollers.

Finally, software averaging of two or more consecutive estimations of R_s and C_s could help with the partial compensation of possible circuit asymmetries.

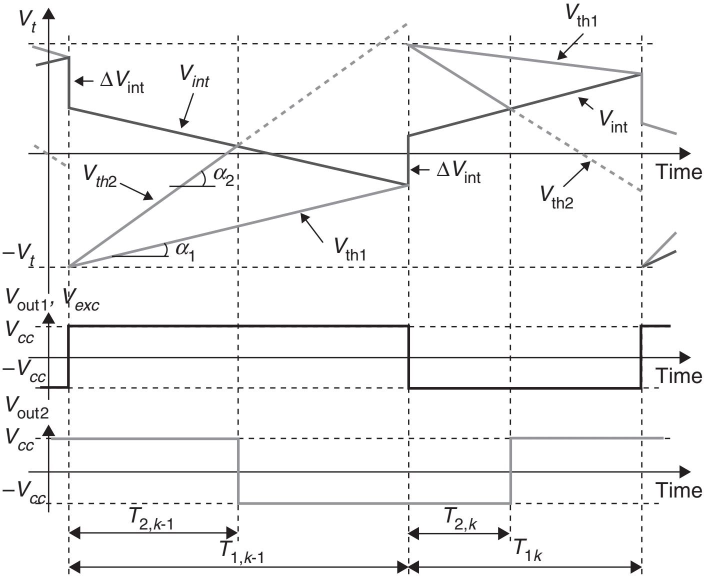
The problem of long measuring times

It is evident from Equation [7.22] that the time needed to perform a measurement is directly proportional to the value of the resistance to be measured. For example, a variation of seven decades in the resistance implies the need to measure time over seven decades, as well. The choice of the circuit parameters allows the designer to shift the time range of the output signal according to the characteristics of the system adopted for the time estimation. From perspective, the strictest requirement is usually the shortest time to be estimated. In fact, if suitable resolution is required in the measurement of time, the clock period of the time measurement system must be much less than the time interval to be estimated. Once the designer has set the minimum time duration, the maximum time duration depends on the resistance range to be measured. For example, if a 10 MHz clock frequency is adopted for the time measurement, a time interval of 10 μs can be estimated with a 1% resolution. If a six-decade resistance variation is considered, the maximum measurement time is about 10 s. If this measurement time is too long for a specific application, the designer can move the time interval range down either by increasing the clock frequency or by decreasing the measurement resolution.



7.28 Modification to the circuit in Fig. 7.24 with the introduction of the moving threshold signals.

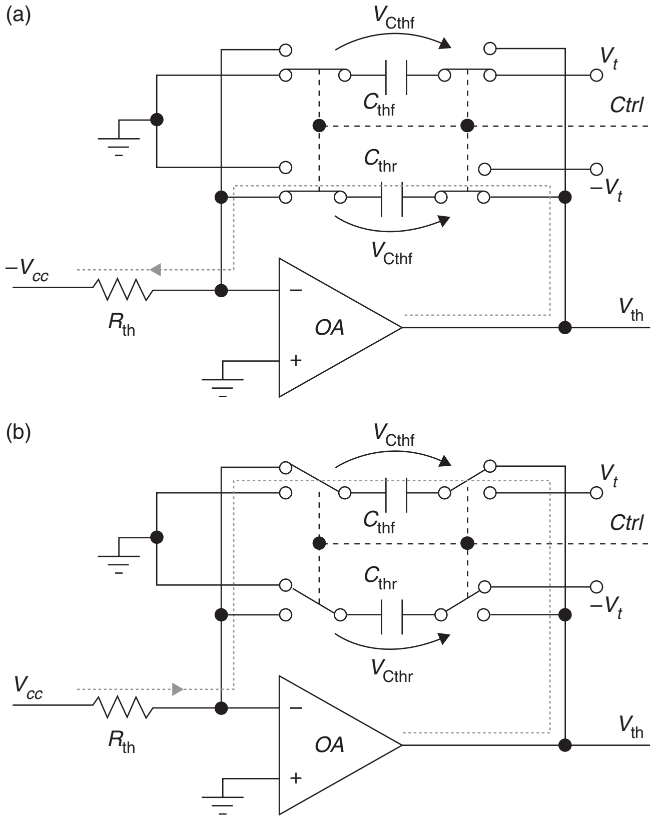
To reduce the measurement time of large resistance values, the oscillator circuit architecture can be modified by introducing the concept of a moving threshold; that is, a threshold signal which is not constant, but moves towards the integrator ramp V_{int} (Depari *et al.*, 2011). Figure 7.28 shows how the circuit in Fig. 7.24 can be modified for this purpose. The former V_{th} signal is now the main threshold V_{th1} and the former fixed threshold (ground voltage) of *Comp*$_2$ is now the second threshold V_{th2}. The generation of the correct threshold signals is performed by the *ThGen*$_1$ and *ThGen*$_2$ blocks, which will be described next. As shown in the time diagram of Fig. 7.29, the threshold signals are designed to be ramps with different slopes (the slope of V_{th2} being steeper than the slope of V_{th1}), but always in the opposite direction with respect to the integrator output ramp V_{int}. As in the previous oscillating circuits, the interception between the main threshold and the integrator output V_{int} determines the circuit commutation, with the consequent change in the direction of the integrator output signal; thus, the threshold signals must change direction as well. To guarantee that the oscillator performs correctly, for every commutation of the circuit the threshold signals must start from established values (V_t and $-V_t$ and in Fig. 7.29), possibly symmetrical with respect to the ground voltage. Moreover, it has to be assured that the integrator output V_{int} is always included in the range between the starting points of the two thresholds V_t and $-V_t$, also when the commutations ΔV_{int} due to the parasitic capacitance happen. As can be seen in the time diagram, even if the integrator output V_{int} has a quasi-flat or even flat slope, the interception with the threshold V_{th1}, thus the circuit commutation, is guaranteed by the movement of the threshold signal. A type of output range compression is applied, which limits the maximum measurement time to $T_{\max}$ as in Equation [7.31], where $|\alpha_1|$ is the absolute value of V_{th1} slope.



7.29 Timing diagram of the oscillator circuit with the moving threshold approach.

$$T_{\max} = \frac{V_t}{|\alpha_1|} \qquad [7.31]$$

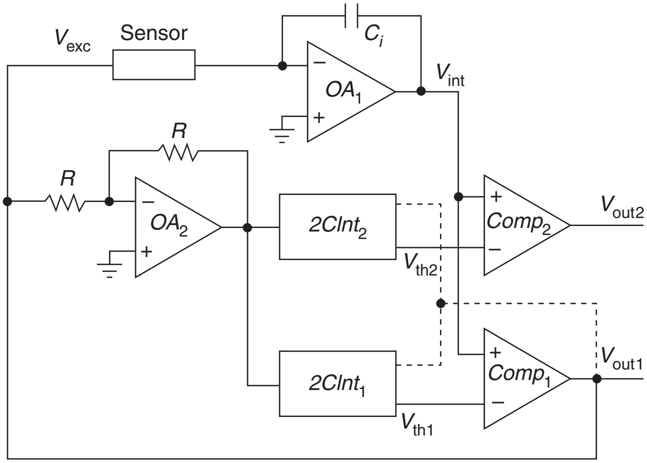
The *ThGen* circuits for the generation of suitable threshold signals can take advantage of the integrator scheme, which allows an output ramp to be generated. The slope and the direction of the output ramp can be easily determined with the integrator parameters and input voltage polarity; conversely, setting the correct starting values for the output ramp, according to the diagram in Fig. 7.29, requires a suitable circuit for the integrator reset. For this purpose, a possible solution is shown in Fig. 7.30, where the two-capacitor integrator (2*CInt*) is illustrated. This integrator scheme includes two integration capacitors; C_{thr} is dedicated to the generation of a rising threshold ramp during a semi-cycle (with the *2CInt* input at $-V_{cc}$, as in Fig. 7.30a), whereas C_{thf} is used for the falling ramp, during the other semi-cycle (with the *2CInt* input at V_{cc},, as in Fig. 7.30b). Thus, only one capacitor at a time is connected to *OA*; the disconnected capacitor is charged to a suitable voltage value, to guarantee the correct starting point for the *OA* output signal for the next semi-cycle. If the two capacitors are chosen with the same value C_{th}, the output ramp has a symmetrical slope, given by Equation [7.32].



7.30 The two-capacitor integrator (*2CInt*) for the moving threshold generation: (a) generation of a rising ramp; (b) generation of a falling ramp.

$$|\alpha| = \frac{V_{cc}}{R_{th} C_{th}} \qquad [7.32]$$

The connection/disconnection of the capacitors is performed by means of a switch network, driven by a signal *Ctrl* which commutates every semi-cycle, such as V_{out1}. In addition, it should be pointed out that, in every semi-cycle, the *2CInt* input signal must have opposite polarity to the input signal V_{exc} of the main integrator, in order to generate a threshold ramp with a direction that is always opposite to V_{int} (see Fig. 7.29). It is clear that, once the polarity has been changed, the output signal V_{out1} can be used as the *2CInt* input signal. Figure 7.31 shows the overall circuit scheme, where *2CInt₁* and *2CInt₂* are two distinct *2CInt* blocks as in Fig. 7.30.



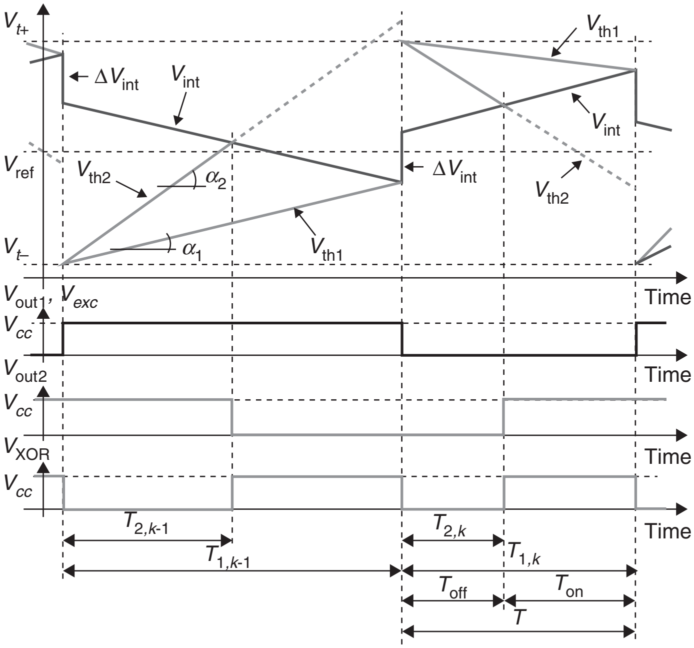
7.31 The oscillating circuit with moving threshold for the limitation of the maximum measuring time. *2CInt* blocks are as in Fig. 7.30.

The switch network is the crucial part of the circuit and thus it should be designed with particular care. Commutation delays need to be as equal as possible among all switches and should be much smaller than the minimum expected time interval to be measured, in order not to have a significant effect on the measurement. The on-state resistance, which is in series with the integrator capacitor, causes an offset to appear on the threshold ramp V_{th}, the value of which depends on the current to be integrated, coming from the integrator input resistance R_{th}. Small values of on-state resistance are advisable to minimize this effect. The off-state resistance should be much larger than the chosen integrator input resistance R_{th}; in this way, currents flowing through open switches do not significantly alter the integration of the current flowing through R_{th} and, thus, the threshold ramp slope.

The estimation of R_s and C_s can be obtained for every semi-cycle k, by measuring the time intervals $T_{1,k}$ and $T_{2,k}$ determined by the output signals V_{out1} and V_{out2} and applying Equations [7.33] and [7.34] where r_α is defined as the ratio between the slope of V_{th1} and V_{th2} ($0 < r_\alpha < 1$) and $V_{\text{exc}} = V_{cc}$.

$$R_s = r_\alpha \frac{T_{\max} V_{\text{exc}}}{2\alpha V_t C_i} \frac{T_{1,k} - T_{2,k}}{T_{2,k} - r_\alpha T_{1,k}} \qquad [7.33]$$

$$C_s = \frac{V_t C_i}{T_{\max} V_{\text{exc}}} \left(T_{\max} - \frac{1 - r_\alpha}{r_\alpha} \frac{T_{1,k} T_{2,k}}{T_{1,k} - T_{2,k}} - T_{1,k-1} \right) \qquad [7.34]$$


V_{t+}
ΔV_{int}
V_{int}
V_{th1}
V_{int}
V_{ref}
V_{th2}
α_2
ΔV_{int}
V_{th2}
V_{t-}
α_1
V_{th1}
V_{out1}, V_{exc}
Time
V_{cc}
V_{out2}
Time
V_{cc}
V_{XOR}
Time
V_{cc}
Time
$T_{2,k-1}$
$T_{1,k-1}$
$T_{2,k}$
$T_{1,k}$
T_{off}
T_{on}
T

7.32 Timing diagram of the oscillating circuit with moving threshold with single power supply and single output signal.

A more reliable estimation can be obtained by averaging the measurement related to two semi-cycles, thus partially compensating for possible circuit asymmetries.

It should be noted that, as in the case of constant thresholds, it is possible to design a single-supply single-output circuit, by shifting the voltage levels to positive values and by adding an XOR logic gate to combine V_{out1} and V_{out2}, as shown in Fig. 7.32. The reference voltage should be placed in the middle of V_{int} range i.e, $V_{\text{ref}} = V_{\text{cc}}/2$; the new starting points of the threshold values should be symmetrical with V_{ref} (in Fig. 7.32, $V_{t+} = V_{\text{ref}} + V_t$ and $V_{t-} = V_{\text{ref}} - V_t$). It is still possible to estimate the value of R_s and C_s in every semi-cycle by using Equations [7.33] and [7.34], with $V_{\text{exc}} = V_{cc}/2$. Conversely, taking advantage of having a single output, the estimation of R_s and C_s can be obtained by the measurement of the period T and the duty-cycle D of V_{XOR} and by using Equations [7.35] and [7.36], where T_{prev} is the period measured in the previous cycle.

$$R_s = r_\alpha \frac{T_{\max} V_{\text{cc}}}{4 V_t C_i} \frac{D-1}{1-r_\alpha - D} \qquad [7.35]$$

$$C_s = \frac{V_t C_i}{T_{\max} V_{\text{exc}}} \left(T_{\max} - \frac{1 - r_\alpha}{r_\alpha} \frac{1 - D}{D} T - T_{\text{prev}} \right) \qquad [7.36]$$

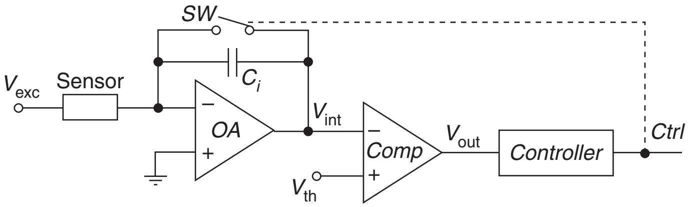
It should be noted that R_s depends only on the duty-cycle D of V_{XOR}, as though it were a pulse width modulation (PWM) signal. This characteristic can be made use of to design simple readout electronics, by low-pass filtering the output signal V_{XOR} and measuring the resulting voltage by means of an A/D converter. This allows the microcontroller of the smart sensor to extract the information about the D (and therefore about R_s) without the need of a time unit hardware or having to implement duty-cycle measurement routines.

Obviously, converting the information to the voltage domain leads to additional problems, such as vulnerability to interference and, thus, increased measurement uncertainty. This kind of approach should be therefore used only when low-cost is more important than system performance and very cheap microcontrollers with limited computational resources need to be used.

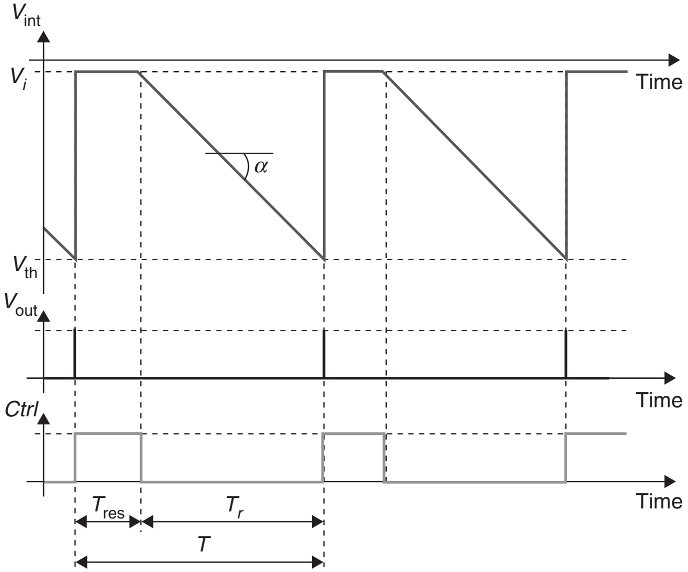
7.4.2 Systems with constant sensor excitation voltage

Key advantages of the oscillating circuits previously described are their simplicity and the opportunity they present, due to their symmetrical architecture, to compensate partially for circuit nonidealities. The main drawback is that the sensor bias voltage V_{exc} is an alternate signal, which is necessary to guarantee the oscillating behaviour by commutating the integrator output voltage V_{int}. Troubles arising from the presence of parasitic capacitive effects can be reduced by using the aforementioned modified solutions; however, the problems related to the variation of the excitation voltage of MOX sensors will still be present.

By using the integrator architecture and a suitable reset circuit, it is possible to perform the R_s estimation only during one phase of the integration (e.g., during the V_{int} falling ramp) and to restart the measurement without the need to modify the sensor excitation voltage V_{exc} (Depari *et al.*, 2006). This concept is illustrated by using the circuit in Fig. 7.33, the timing diagram of which is presented in Fig. 7.34. Starting from the basic integrator circuit, a reset switch *SW*, driven by the control signal *Ctrl*, is placed in parallel to the integrator capacitance C_i. When activated (closed), *SW* forces the integrator output V_{int} to the initial value V_i, close to the ground voltage (component nonidealities make V_i differ from the ideal value – i.e., ground voltage). In the example in Fig. 7.33, V_{exc} is a positive and constant voltage and thus, under the hypothesis of constant R_s, V_{int}, behaves as a falling ramp when *SW* is open. A negative threshold voltage V_{th} and a comparator *Comp*


SW
C_i
V_{exc}
Sensor
OA
V_{int}
Comp
V_{th}
V_{out}
Controller
Ctrl

7.33 The basic circuit for the RTC implementation with constant sensor excitation voltage.


V_{int}
V_i
Time
α
V_{th}
V_{out}
Time
Ctrl
Time
T_{res}
T_r
T

7.34 Timing diagram of the circuit in Fig. 7.33.

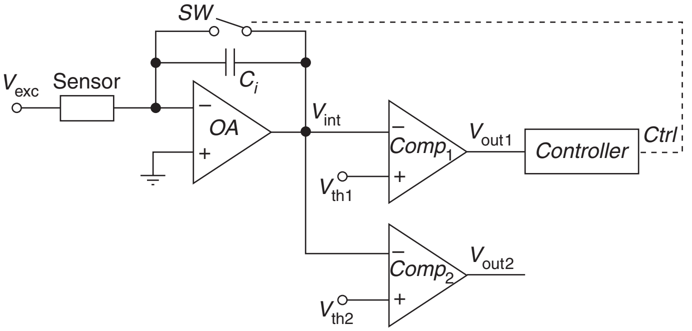
are used to generate a pulsed signal V_{out}. The time T_r taken by the V_{int} to intercept V_{th} is related to R_s by means of Equation [7.37].

$$T_r = \frac{|V_{th} - V_i|}{\alpha} = \frac{|V_{th} - V_i|}{V_{exc}} R_s C_i \tag{7.37}$$

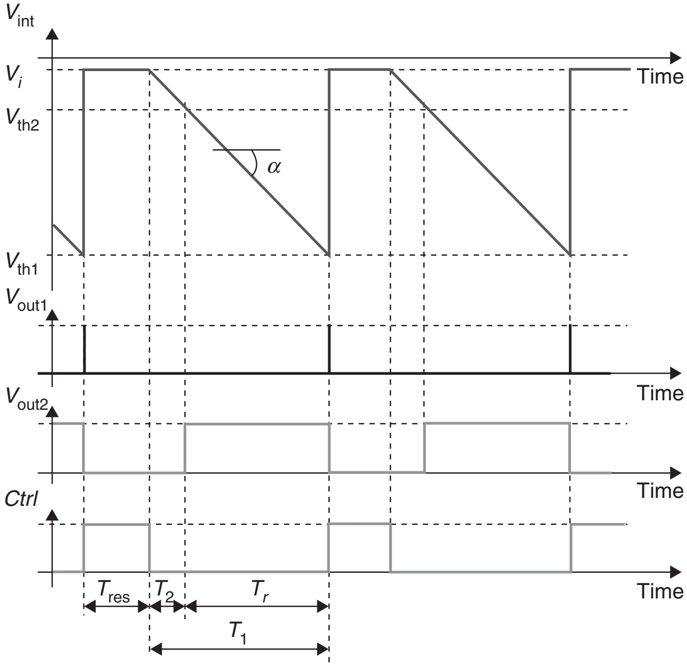
The *Controller* stage in Fig. 7.33 is devoted to generate the switch control signal *Ctrl*. This can be realized with a simple monostable circuit, triggered by the comparator output signal V_{out}, as well as directly implemented in the microcontroller or PLD of the smart sensor. The time duration T_{res} of the reset phase should be chosen in order to guarantee a complete reset of the integrator. As previously stated, component nonidealities, especially those related to the operational amplifier *OA* and the switch *SW*, cause the initial value V_i not to be zero; even if Equation [7.37] considers this effect, the evaluation of V_i is not easy and it can be affected by several uncertainty factors, such as temperature. The work in (Depari *et al.*, 2006) proposes that a feedback circuit be applied in parallel to *SW* and controlled by *Controller*, in order to reduce the less than ideal reset of the integrator; thus, it is possible to consider $V_i \approx 0$ V. The comparator output V_{out} is a narrow pulsed signal; in fact, as soon as it commutates to the high level, the *Controller* stage issues the reset command, making the integrator output V_{int} increase rapidly and, thus, V_{out} to commutate to the low level once more. From the measurement of the period T of V_{out} or *Ctrl*, it is possible to estimate the time T_r (by subtracting T_{res}), and to calculate the R_s value from Equation [7.37].

A different approach to solve the problem related to V_i is to use two threshold values (V_{th1} and V_{th2}, with $|V_{\text{th2}}| < |V_{\text{th1}}|$) and to consider the V_{int} ramp behaviour only between such thresholds. Figures 7.35 and 7.36 show the circuit schematic and the timing diagram related to this solution. As in the previous case, the *Ctrl* signal is issued by the *Controller* block when triggered by the comparator output signal related to the main threshold voltage (V_{th1}). The time interval T_r is defined as the time taken by V_{int} to span between the two thresholds; it can be obtained by subtracting the time interval T_2 (time between the release of the reset and the interception of V_{in} with the threshold values V_{th1} and V_{th2}, respectively: see Fig. 7.36) from T_1, as well as by estimating the high level time of the comparator output *Comp₂*. The relationship between the sensor resistance R_s and T_r is obtained from Equation [7.37], where V_{th} is now V_{th1} and V_{th2} replaces V_i, as in Equation [7.38].

$$T_r = T_1 - T_2 = \frac{|V_{\text{th1}} - V_{\text{th2}}|}{\alpha} = \frac{|V_{\text{th1}} - V_{\text{th2}}|}{V_{\text{exc}}} R_s C_i \qquad [7.38]$$



7.35 The basic circuit for the RTC implementation with constant sensor excitation voltage; solution with two threshold voltages.

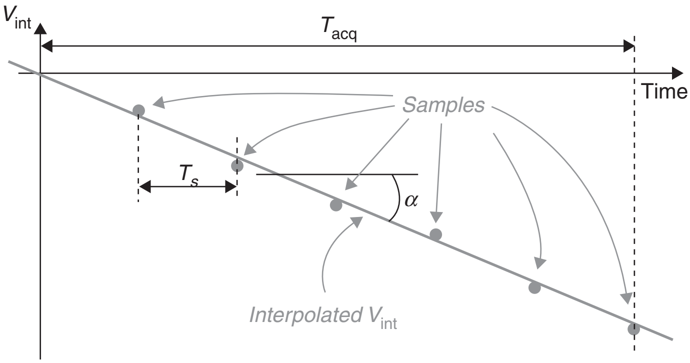


7.36 Timing diagram of the circuit in Fig. 7.35.

Note that the initial voltage V_i of the integrator output no longer affects the estimation of R_s. This is obtained because now the integrator output ramp is evaluated between two known voltages (V_{th1} and V_{th2}); this solution, therefore, can avoid the use of the feedback circuit in parallel with *SW*, thus keeping the circuit topology quite simple. However, whereas in the previous case a simple period estimation procedure was needed, in this case a slightly more complex time measurement unit or routine needs to be implemented.

Long measuring time problem

It is evident from Equations [7.37] and [7.38] that the ramp time T_r is proportional to the value of the resistance R_s to be estimated, as happens with the oscillator-based circuit of Fig. 7.21. The same considerations regarding the choice of the circuit parameters to adapt the output signal timings to the sensor resistance range and the characteristics of the time estimation unit can be applied. As stated previously, when a wide range of resistances needs to be estimated, a long measurement time usually occurs in the upper part of the resistive range; this could be a problem for particular sensor applications. Even if, as in the case of the oscillating circuits, a moving threshold approach can be followed, the architecture of the constant sensor excitation



7.37 Concept of the LMS interpolation of the integrator output ramp from the acquisition of V_{int} samples.

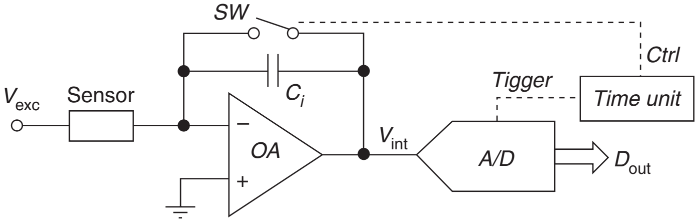
voltage suggests adopting a more effective solution for shortening the measurement time.

The key point for the R_s estimation is the determination of the slope α of V_{int}; this goal can be achieved by acquiring a suitable number N of samples of the V_{int} signal with an A/D converter, as shown in Fig. 7.37, and by applying the LMS interpolation algorithm. The sample rate F_s, determines the time distance T_s between samples ($T_s = 1/F_s$); once all the N samples are collected, the measurement process can be either ended or iterated. The resulting acquisition time can be obtained as $T_{acq} = N \cdot T_s$. It should be noted that T_{acq} is independent of the R_s value and the measurement time is constant; this characteristic is important in applications in which a certain degree of synchronization between the sensor data acquisition and other operations must be achieved (Depari *et al.*, 2012b).

As stated previously, the estimation of α is accomplished by means of the LMS algorithm; that is by applying Equation [7.39], where $V_{int,i}$ is the sample i value, and $T_{s,i}$ is the time distance of the sample i from the beginning of the acquisition ($t = 0$ in Fig. 7.37).

$$|\alpha| = \left| \frac{\sum_{i=1}^{N} (T_{s,i} - \overline{T_s})(V_{int,i} - \overline{V_{int}})}{\sum_{i=1}^{N} (T_{s,i} - \overline{T_s})^2} \right| \quad \text{with} \quad \overline{T_s} = \frac{\sum_{i=1}^{N} T_{s,i}}{N} \quad \text{and} \quad \overline{V_{int}} = \frac{\sum_{i=1}^{N} V_{int,i}}{N} \tag{7.39}$$

To apply this technique, a circuit such as that in Fig. 7.38 has to be implemented. The *Time Unit* block is devoted to the generation of the *Trigger* signal to the A/D converter, in order to acquire the V_{int} samples with the correct timings. In addition, it has to drive the reset switch *SW* to start and end the



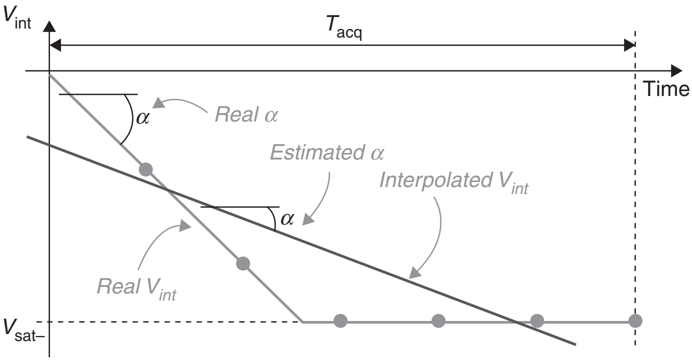
7.38 Simple circuit for the R_s estimation by using the LMS estimation approach.

estimation. The converted samples D_{out} have to be acquired and elaborated by the microcontroller or PLD of the smart sensor, which has to be able to implement the estimation by applying Equation [7.39] and the first part of Equation [7.21]. Note that the signal V_{int} varies during the sample process, thus a sample-and-hold operation needs to be performed before the conversion. Alternatively, the conversion time of the A/D converter should be as short as possible, in order to consider the V_{int} reasonably constant during the conversion itself. This can be easily achieved since, as will be shown next, the method is usually applied for large resistances values and thus slow variable V_{int}. Moreover, it is worth noting that, for signal reconstruction, using the calculation of α by means of Equation [7.39], *a priori* knowledge of the monotonous nature of the sampled signal V_{int} is employed. The only constraint for the correct signal reconstruction is that the number of samples N is at least two (due to the presence of noise, a number of samples greater than ten is advised, to guarantee the method reliability). This fact allows aliasing problems to be ignored and, therefore, the usual anti-aliasing low-pass filter on the signal to sample can be omitted.

The operating range of this kind of solution is limited at the lower part by the saturation voltage of *OA*. In fact, as visible in Fig. 7.39, if α is too large (small R_s value) the integrator output V_{int} can reach the negative saturation voltage V_{sat-} before the last sample has been taken. In this case, the estimation of α by means of the LMS algorithm is not correct. The maximum value α_{max} and the related minimum value $R_{s,min}$ of the sensor resistance are given by Equations [7.40] and [7.41], respectively.

$$\alpha_{\max} = \frac{V_{\text{sat}-}}{NT_s} \qquad [7.40]$$

$$R_{s,\min} = \frac{V_{\text{exc}}}{|V_{\text{sat}-}|}\frac{NT_s}{C_t} \qquad [7.41]$$



7.39 The problem of the LMS interpolation in the event of reaching the saturation voltage of the operational amplifier OA.

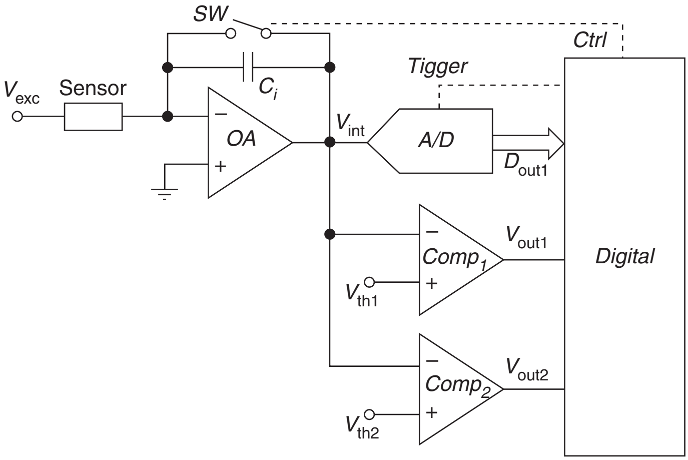
Conversely, the limit at the upper part of the range mainly depends on the A/D converter characteristics and, more in particular, by the effective number of bits (ENOB). In fact, when R_s is very large and, therefore, V_{int} almost flat (very small α value), the total variation of V_{int} within the acquisition time T_{acq} can be less than the minimum quantity detectable by the A/D converter. In addition, the noise affecting the integrator output V_{int} makes the LMS interpolation, obtained with samples acquired in such conditions, totally unreliable. In greater detail, considering $V_{\text{A/D}}$ and B as the input range and the ENOB of the A/D converter, the minimum value $\alpha_{\min}$ and the related maximum value $R_{\text{s,max}}$ of the sensor resistance are given by Equations [7.42] and [7.43], respectively.

$$\alpha_{\min} = \frac{V_{A/D}}{2^B N T_s} \qquad [7.42]$$

$$R_{s,\max} = \frac{V_{\text{exc}}}{V_{A/D}} \frac{2^B N T_s}{C_i} \qquad [7.43]$$

It is worth noting that, if the circuit parameters are chosen in order to have $V_{\text{A/D}} = |V_{\text{sat}-}|$, then the ratio between the maximum and minimum R_s value is determined by the ENOB B, and, in particular, $R_{\text{s,max}} = 2^B \cdot R_{\text{s,min}}$. In a practical example, if the resistance estimation needs to be performed over four decades, an A/D converter with $B = 14$ bits should be used; however, if the range needs to be expanded to six decades, then $B = 20$ bits, causing a significant increase in the cost of the A/D converter.

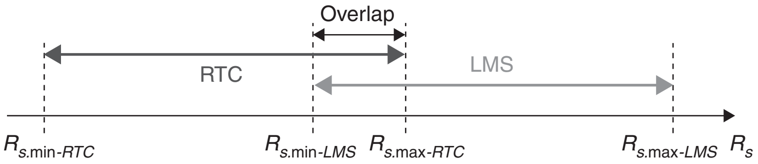
Due to the similarity of the scheme in Figs 7.35 and 7.38, an expansion of the operating range of the front-end can be obtained by combining both



7.40 The circuit uniting the circuits in Figs 7.35 and 7.38 for the expansion of the R_s estimation range.

the RTC and LMS approaches and suitably designing the circuit parameters to dedicate an estimation method to a specific R_s sub-range (Depari *et al.*, 2012a). The resulting circuit is shown in Fig. 7.40. The block *Digital* incorporates the functions of the previous stages *Controller* of Fig. 7.35 and *Time Unit* of Fig. 7.38; moreover, a decision about which method is able to furnish a correct estimation for a particular R_s value must be performed by this block. Due to the complexity of the operations to be handled by *Digital*, it is advisable that such a block is implemented directly in the microcontroller or PLD of the smart sensor. In this way, *Digital* can be also devoted to the measurement of the time interval T_r (for RTC estimation, see Fig. 7.36), and the acquisition of the V_{int} samples (for an LMS estimation, see Fig. 7.37).

The basic idea for the R_s range expansion is that the RTC method can be used as long as it provides a correct estimation within the desired measurement time. If the threshold value V_{th1} cannot be reached by V_{int}, because R_s is too great (see Fig. 7.36), then the LMS approach is applied. Analogously, the LMS estimation can be used as long as the ramp V_{int} does not reach the *OA* saturation voltage; if this happens, because of R_s is too small (see Fig. 7.39), then the RTC approach is applied. Summarizing, the RTC approach is suitable for use in the lower part of the measurement range, whereas the LMS technique can be applied when R_s is in the upper part. A partial overlap of the two methods in the mid-range can be useful for circuit calibration. Figure 7.41 illustrates the R_s range partition previously discussed. The overall minimum resistance $R_{\text{s,min}}$ which can be estimated is determined by the RTC method ($R_{\text{s,min-RTC}}$) and, as previously described, it



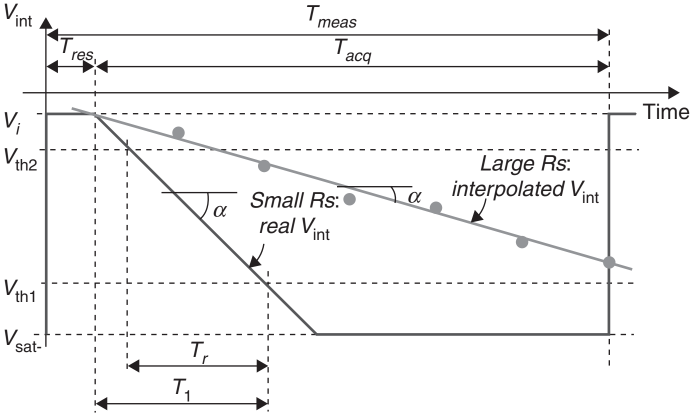
7.41 Sensor resistance range partition in the event of using the circuit in Fig. 7.40, uniting the RTC and the LMS approaches.

depends on the desired resolution of the time measurement. Conversely, the maximum resistance $R_{s,max}$ which can be estimated is determined by the LMS method ($R_{s,max\text{-}LMS}$) and, as detailed in this section, it mainly depends on the ENOB of the A/D converter. To guarantee a continuity in the two sub-ranges, the circuit parameters must be chosen in order to have the lower limit of the LMS approach ($R_{s,max\text{-}LMS}$) less than the upper limit of the RTC method ($R_{s,max\text{-}RTC}$). Figure 7.42 reports the time diagram of the two working modes. If the time T_1 (time needed by V_{int} to intercept the threshold voltage V_{th1}) is longer than the acquisition time T_{acq}, only the LMS method can be applied. If V_{int} crosses the voltage threshold before T_{acq} expires, then the RTC approach is appropriate. In the case of the RTC approach, if V_{int} do not reach the saturation voltage V_{sat-}, the LMS method is valid as well, yielding the aforementioned overlap region. The overall measurement time T_{meas} is given by the sum of the acquisition time T_{acq} and the reset time T_{res} and it is constant, independent of the R_s value. Usually, the time T_{res} is chosen to be much less than T_{acq}; therefore, $T_{meas} \approx T_{acq}$. A practical implementation of this method is detailed in (Depari *et al.*, 2012a).

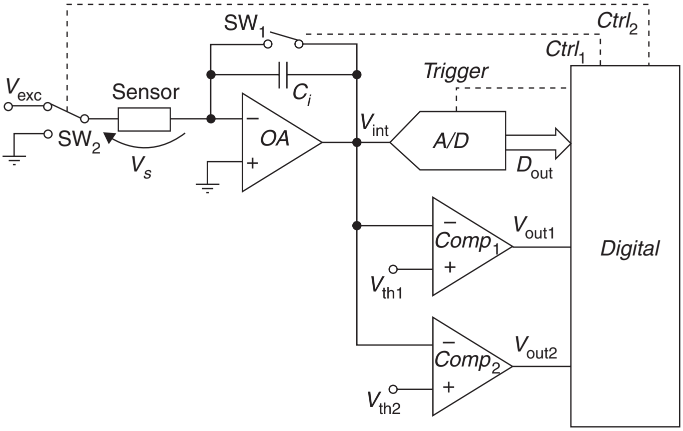
Parasitic capacitance estimation

The main advantage of the constant sensor bias voltage approach is that the parasitic capacitance of the sensor is not excited and therefore does not influence the R_s estimation. As stated before, the parasitic capacitance can arise from the sensor itself, but it can also include effects due to long connections between the sensor and the front-end. In particular, the monitoring of the parasitic capacitive effects, described by the parasitic capacitor C_s in parallel with R_s, could help the smart sensor with diagnostic operations, such as the control of sensor connection integrity.

The estimation of C_s can be obtained only if the sensor excitation voltage is varied; in the circuit that combines the RTC and LMS approaches (Fig. 7.41), this can be easily obtained by commutating the sensor excitation voltage to ground voltage during the integrator reset phase (Depari *et al.*, 2012a). In Fig. 7.43, the circuit of Fig. 7.40 has been modified for this

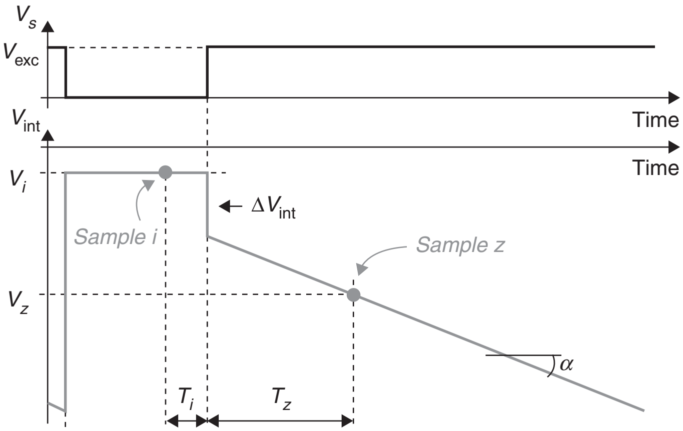


7.42 The time diagram of the circuit in Fig. 7.40 with two different situations: small R_s (RTC method applied) and large R_s (LMS method applied).


SW_1
$Ctrl_2$
$Ctrl_1$
C_i
Trigger
V_{exc}
Sensor
SW_2
V_s
OA
V_{int}
A/D
D_{out}
$Comp_1$
V_{out1}
V_{th1}
Digital
$Comp_2$
V_{out2}
V_{th2}

7.43 Modification of the circuit in Fig. 7.40 for the estimation of the parasitic capacitive effects.

purpose. The sensor excitation voltage V_s can be either the constant voltage V_{exc} or ground voltage, chosen by means of the switch SW_2. The SW_2 control signal $Ctrl_2$ is suitably driven by the *Digital* block. As stated before, V_s needs to be driven to ground voltage during the reset phase, or otherwise to V_{exc}. The same control signal $Ctrl_1$ as for the reset switch SW_1 (the former *Ctrl* and *SW* in Fig. 7.40, respectively) could therefore be used as $Ctrl_2$. On the other hand, if the capacitive effect estimation does not need to be performed during every measurement cycle but, instead, at intervals,



7.44 Time diagram of the circuit in Fig. 7.43 highlighting the parasitic capacitance estimation procedure.

the choice of splitting the two switch control lines affords higher system flexibility.

The time diagram of the circuit in Fig. 7.43 around the reset phase is shown in Fig. 7.44. During the reset phase, the commutation of V_s from V_{exc} to ground voltage does not affect V_{int}, since V_{int} is driven to V_i (close to ground voltage) by the reset switch SW_1. When the reset phase ends, the switch SW_1 is opened and simultaneously the switch SW_2 is driven to connect the sensor to V_{exc}. The sudden variation of the sensor bias voltage creates the charge transfer effect across C_s and C_i, which can be seen in Fig. 7.44 with a step variation ΔV_{int} of the integrator output, the value of which is given by:

$$|\Delta V_{int}| = |\Delta V_s| \frac{C_s}{C_i} = V_{exc} \frac{C_s}{C_i} \qquad [7.44]$$

To evaluate ΔV_{int}, two additional samples (sample *i* and *z*) of V_i can be taken – before a time T_i and after a time T_z with respect to the end of the reset phase. The sample *i* is used to estimate the V_i voltage, which is the starting value of the V_{int} ramp or of the ΔV_{int} step in the event that parasitic capacitance is present. Even if, as in Fig. 7.44, V_{int} is considered constant within the reset phase, the nonidealities of components can cause V_{int} to vary; for instance, to demonstrate an exponential behaviour toward V_i. For this reason, the time T_i needs to be as close as possible to the end of the reset phase. On the other hand, sample *z* accounts for the final value of the ΔV_{int} step. In Fig. 7.44, ΔV_{int} is considered to be an instantaneous voltage variation; component nonidealities – in particular, the finite slew rate of the *OA* – imply that the ΔV_{int} variation has a finite duration. The time T_z therefore needs to be apropriately far from the end of the reset phase, to guarantee the completion of the charge transfer effect.

Note that if the voltage step ΔV_{int} is small enough either not to intercept the threshold voltage V_{th2} (see Fig. 7.42) or not to cause V_{int} to reach the saturation voltage of *OA* (see Fig. 7.39), then either the RTC or LMS method can be used to estimate the V_{int} slope α.

The final value of the voltage step ΔV_{int} can be extrapolated by using the V_{int} value V_z obtained with the sample z and α; the ΔV_{int} can be therefore obtained:

$$|\Delta V_{\text{int}}| = |V_z - V_i| - |\alpha| T_z \quad [7.45]$$

Finally, from Equation [7.44], the value of the parasitic capacitance C_s can be estimated.

7.4.3 Industrial-related aspects

The interface solutions that have been presented can be used for a broad range of applications. However, they are particularly advantageous when dealing with wide-range devices, such as chemical sensors for gas detection. In industrial environment, applications are generally related to the monitoring of environmental quality or product quality; for example, in the detection of food degradation.

In the case of environmental quality monitoring, usually the main requirements are the compactness of the sensor system and a low cost. In fact, these systems are usually placed in different areas of the site to be monitored, in order to guarantee a complete coverage of the places of interest. Moreover, constraints related to sensor system location could require battery-operated devices, thus low power consumption becomes a necessity.

Oscillator-based approaches with constant threshold voltages are ideal for this purpose, due to the simple architecture which helps in keeping the overall system cost low and makes it possible to implement as an integrated circuit, assuring compactness of the system (Ferri *et al.*, 2009). Moreover, as can be seen in Equations [7.26] and [7.27], the circuit parameters appear as multiplicative factors in the sensor estimation relationship; thus, simple calibration procedures are suitable to compensate for circuit nonidealities. Realization of the sensor interface together with a microcontroller or PLD makes it possible to implement data filtering, pre-processing and simple procedures of self-recalibration – thus assuring compensation for instability or aging effects of the circuit components.

With regard to power consumption, best practice is the utilization of the sensors in a pulsed (sleep mode) regime; that is, keeping the sensor system off for a long time and activating it only when taking a measurement. In this case, the sensor readout time plays an important role, because it determines the activation time of the system and, therefore, the power requirement. Short measuring time interfaces, such as the oscillator-based approaches with variable threshold voltages, are better choices for this kind of application. However, it should be mentioned that the price to be paid in this case is not only in the more complex system architecture, but also in a more difficult calibration procedure. In fact, as can be seen from Equations [7.33] and [7.34], the relationships to apply are more complex and a simple calibration/linearization procedure is not suitable for compensating all the circuit nonidealities.

It is worth noting that, in many applications, systems for environmental quality monitoring do not need to be precisely accurate, because they simply have to generate alerts in case the quantity under consideration exceeds a certain threshold level. This level can be suitably lowered to allow for system inaccuracy, still guaranteeing an adequate level of safety, even if false alarms could be generated. In this case, a trade-off between the cost (included the calibration) and the system performance can usually be achieved, with the cost being the greater consideration.

The reverse is relevant for the realization of systems for food quality monitoring: the primary aspect to consider is the system performance; the measurements need to be precise in order to guarantee the effectiveness of the elaboration algorithms applied for the detection of the substances of interest. In these systems, also called *artificial olfactory systems* (AOS), the compactness and the cost are therefore secondary aspects (EOS AROMA catalogue, available at SACMI website, http://www.sacmi.it/default.aspx?ln=en-US). Also, in these applications the measuring time plays a significant role, because advanced elaboration techniques need to work with the dynamic response of the sensors rather than with a steady-state response. A high sampling rate also allows sensor signal oversampling to be performed, thus allowing the microcontroller or PLD of the smart sensor the opportunity to employ more effective data filtering. In addition, it is worth noting that a short measuring time is also beneficial for speeding up the system calibration and the process of collecting data for the training of the AOS. Therefore, even if more complex than the oscillator-based systems, solutions such as that presented in Fig. 7.40 seem to be more suitable for this purpose.

7.5 Conclusion and future trends

The rapid progress in microelectronics and the consequent drop in the cost of electronic components is a key factor for the improvement of sensor systems in general. The availability of low-cost microcontrollers and PLDs with relatively high computational resources allows designers to implement a more complex data acquisition solution. With regard to smart sensor realization, the most suitable and effective calibration, estimation and data elaboration techniques for a specific sensor/application can be therefore chosen and utilized without increasing the overall system cost to an appreciable degree.

Concerning resistive sensors, technical innovations to improve the effectiveness of such devices are continuously in progress. Physical effects, which are the basis of the working principle of resistive sensors, are nowadays well-known and consolidated, and the current improvement of the devices mainly concerns innovative technological solutions. In particular, the research on new materials and their application in the sensing field seems to be the key point for the development of new resistive sensors. The increase of sensitivity towards the physical quantity of interest is a common objective for each kind of sensor.

An exception to this innovation trend of resistive sensors is related to position detection systems realized with potentiometers. At least for linear position detection, magnetostrictive sensors represent a good alternative to potentiometers, due to the contactless characteristic of the moving slide which guarantees better performance in terms of life span, even if at a higher cost (Seco *et al*., 2005).

Regarding new materials for resistive gas sensors, issues related to the improvement of sensor selectivity to a specific target substance or group of substances is one of the main concerns (Zhu *et al*., 2011). To this end, particular sensor operating modes, in terms of pulsed thermal excitation and multivariate data analysis, are also being explored and they are proving to be quite satisfactory (Bicelli *et al*., 2007; Huang *et al*., 2006; Ponzoni *et al*., 2012).

Another key point for the improvement of resistive gas sensors is the choice of materials and operating techniques that can show good sensitivity when operating at low temperatures, ideally at the temperature of the environment (Sharmaa *et al*., 2011). This aspect is particularly important when power consumption is the main concern (e.g., as with battery-operated smart sensors), since the power required by the sensor heating operation is definitely the greatest draw on power, unlike that needed for sensor excitation and data readout/elaboration. Finally, innovative technologies in the field of nanostructures are under examination for the realization of sensors with improved sensitivity and reduced power consumption, because of optimization of the sensor sensitive area and the optimization of sensor size (Sberveglieri *et al.*, 2007).

In order to be suitable for the new sensor requirements, electronic interfaces usually tend to follow the evolution of sensors. For example, for pulse-operated sensors, solutions with a short measuring time and synchronized thermal excitation management have been proposed (Depari *et al.*, 2012b). Realization of the electronic sensor interface as an integrated circuit is also a key point (Ferri *et al.*, 2009) as this is in keeping with the present trend for sensor miniaturization (Jiang *et al.*, 2007). Advantages that can be obtained with interface integration, besides the reduction in system size, can also be found in improved circuit reliability and stability, and reduced power consumption.

Due to the progressively higher computational resources available in inexpensive devices, all modern sensor interfaces include a microcontroller or PLD. They are used for the proper management of interfaces, but also provide advanced functionalities such as the opportunity to apply sensor compensation, data filtering and elaboration and so on, thus delivering the concept of the smart sensor. Lately, particular effort has been put into facilitating data exchange of smart sensors with existing infrastructures, by using typical ICT technologies, such as Bluetooth or USB connections to realize networks of sensors (Depari *et al.*, 2007; Zhang and Xiao, 2009). Even if still at the research stage, the trend in smart sensors for industrial applications is also directed to the implementation of sensor networks, using the modern and promising communication technologies of the industrial world, such as real time Ethernet (RTE) and WirelessHART.

7.6 References

Bicelli, S., Depari, A., Faglia, G., Flammini, A., Fort, A., Mugnaini, M., Ponzoni, A., Vignoli, V. and Rocchi, S. (2007) ‘Model and experimental characterization of dynamic behavior of low-power carbon monoxide MOX sensors operated with pulsed temperature profiles’, *IEEE Trans. Instrum. Meas.*, **58**, 1324–1332.

De Marcellis, A., Depari, A., Ferri, G., Flammini, A., Marioli, D., Stornelli, V. and Taroni, A. (2008) ‘Uncalibrated integrable wide-range single-supply portable interface for resistance and parasitic capacitance determination’, *Sens. Act. B: Chem.*, **132**, 477–484.

Depari, A., De Marcellis, A., Ferri, G. and Flammini, A. (2011) ‘A complementary metal oxide semiconductor–integrable conditioning circuit for resistive chemical sensor management’, *IOP Meas. Sci. Tech.*, **22**, 1–7.

Depari, A., Ferrari, P., Flammini, A., Marioli, D. and Taroni, A. (2007) 'A VHDL model of a IEEE1451.2 smart sensor: characterization and applications', *IEEE Sens. J.*, **7**, 619–626.

Depari, A., Flammini, A., Marioli, D., Rosa, S. and Taroni, A. (2006) 'A low-cost circuit for high-value resistive sensors varying over a wide range', *IOP Meas. Sci. Tech.*, **17**, 353–358.

Depari, A., Flammini, A., Marioli, D., Sisinni, E., Comini, E. and Ponzoni, A. (2012a) 'An electronic system to heat MOX sensors with synchronized and programmable thermal profiles', *IEEE Trans. Instrum. Meas.*, **61**, 2374–2383.

Depari, A., Flammini, A. and Sisinni, E. (2012b) 'Measurement of resistance and capacitance of MOX sensors with high sampling rate', *IEEE Trans. Instrum. Meas.*, **61**, 2483–2491.

Faglia, G., Comini, E., Cristalli, A., Sberveglieri, G. and Dori, L. (1999) 'Very low power consumption micromachined CO sensors', *Sens. Act. B: Chem.*, **55**, 140–146.

Ferri, G., De Marcellis, A., Di Carlo, C., Stornelli, V., Flammini, A., Depari, A. and Nader, J. (2009) 'A single-chip integrated interfacing circuit for wide-range resistive gas sensor arrays', *Sens. Act. B: Chem.*, **143**, 218–225.

Ferri, G., Stornelli, V., De Marcellis, A., Flammini, A. and Depari, A. (2008) 'Novel CMOS fully integrable interface for wide-range resistive sensor arrays with parasitic capacitance estimation', *Sens. Act. B: Chem.*, **130**, 207–215.

Flammini, A., Marioli, D. and Taroni, A. (2004) 'A low-cost interface to high value resistive sensors varying over a wide range', *IEEE Trans. Instrum. Meas.*, **53**, 1052–1056.

Grassi, M., Malcovati, P. and Baschirotto, A. (2005) 'A high-precision wide-range front-end for resistive gas sensors arrays', *Sens. Act. B: Chem.*, **111–112**, 281–285.

Huang, J.R., Li, G.Y., Huang, Z.Y., Huang, X.J. and Liu, J.H. (2006) 'Temperature modulation and artificial neural network evaluation for improving the CO selectivity of SnO_2 gas sensor', *Sens. Act. B: Chem.*, **114**, 1059–1063.

Jiang, W.F., Xiao, S.H., Feng, C.Y., Li, H.Y. and Li, X.J. (2007) 'Resistive humidity sensitivity of arrayed multi-wall carbon nanotube nests grown on arrayed nanoporous silicon pillars', *Sens. Act. B: Chem.*, **125**, 651–655.

Lundstrom, I. (1996) 'Approaches and mechanisms to solid state based sensing', *Sens. Act. B: Chem.*, **35**, 11–19.

Meixner, H. and Lampe, U. (1996) 'Metal oxide sensors', *Sens. Act. B: Chem.*, **33**, 198–202.

Olthuis, W., Sprenkels, A.J., Bomer, J.G. and Bergveld, P. (1997) 'Planar interdigitated electrolyte-conductivity sensors on an insulating substrate covered with Ta_2O_5', *Sens. Act. B: Chem.*, **43**, 211–216.

Ponzoni, A., Depari, A., Comini, E., Faglia, G., Flammini, A. and Sberveglieri, G. (2012) 'Exploitation of a low-cost electronic system, designed for low-conductance and wide-range measurements, to control metal oxide gas sensors with temperature profile protocols', *Sens. Act. B: Chem.*, **175**, 149–156.

Sauerwald, T., Skiera, D. and Kohl, D. (2005) 'Field induced polarisation and relaxation of tungsten oxide thick films', *Thin Solid Films*, **490**, 86–93.

Sberveglieri, G., Baratto, C., Comini, E., Faglia, G., Ferroni, M., Ponzoni, A. and Vomiero, A. (2007) 'Synthesis and characterization of semiconducting nanowires for gas sensing', *Sens. Act. B: Chem.*, **121**, 208–213.

Sberveglieri, G., Comini, E., Faglia, G., Atashbar, M. Z. and Wlodarski, W. (2000) 'Titanium dioxide thin films prepared for alcohol microsensor applications', *Sens. Act. B: Chem.*, **66**, 139–141.

Seco, F., Martín, J.M., Jiménez, A.R. and Calderón, L. (2005) 'A high accuracy magnetostrictivelinearpositionsensor', *Sens. Act. A: Phys.*, **123–124**, 216–223.

Sharmaa, A., Tomarb, M. and Gupta, V. (2011) 'SnO_2 thin film sensor with enhanced response for NO_2 gas at lower temperatures', *Sens. Act. B: Chem.*, **156**, 746–752.

Zhang, Y. and Xiao, H. (2009) 'Bluetooth-based sensor networks for remotely monitoring the physiological signals of a patient', *IEEE Trans. Inf. Tech. in Biomed.*, **13**, 1040–1048.

Zhu, L.F., She, J.C., Luo, J.Y., Deng, S.Z., Chen, J., Ji, X.W. and Xu, N.S. (2011) 'Self-heated hydrogen gas sensors based on Pt-coated $W_{18}O_{49}$ nanowire networks with high sensitivity, good selectivity and low power consumption', *Sens. Act. B: Chem.*, **153**, 354–360.

8

Reconfigurable ultrasonic smart sensor platform for nondestructive evaluation and imaging applications

J. SANIIE, E. ORUKLU and S. GILLILAND, Illinois Institute of Technology, USA and S. ASLAN, Texas State University, USA

DOI: 10.1533/9780857099297.1.205

Abstract: This chapter presents a reconfigurable ultrasonic smart sensor platform (RUSSP) for real-time signal analysis and image processing. The RUSSP is designed to facilitate the development and implementation of ultrasonic signal processing algorithms with embedded software and reconfigurable hardware. It provides the opportunity to explore the full design space including software only, hardware only and hardware/software co-design architectures. Multiple case studies are investigated including real-time ultrasonic flaw detection, signal decomposition and parameter estimation on the RUSSP, which hosts a Linux operating system and reconfigurable fabric.

Key words: ultrasonic signal processing, system-on-chip, nondestructive testing, flaw detection, field programmable gate arrays (FPGAs), reconfigurable computing.

8.1 Introduction

In recent years, high-speed signal analysis has been used to modernize methods of ultrasonic measurement, imaging and testing (Lu *et al*., 2006; Goldsmith *et al*., 2008; Kunita *et al*., 2008; Oruklu and Saniie 2009; Weber, *et al*., 2011). Ultrasonic technologies have continued to gain in popularity due to their high precision, low cost and shrinking footprint of the sensors. It is now possible to measure object properties and produce high resolution images and video that provide a complete three-dimensional view of the internal physical geometrical characteristics within an object. These capabilities have given new insight into fields such as structural health monitoring, material science and medical imaging.

This chapter presents the design and application of reconfigurable ultrasonic smart sensor platform (RUSSP) based on field programmable gate arrays (FPGAs). The proposed system provides real-time signal processing for nondestructive evaluation (NDE) and imaging applications using ultrasonic sensors ranging from 20 kHz to 20 MHz operational frequencies. Specifically, hardware implementations suitable for portable systems with networking capability are discussed. In practice, NDE of materials requires the design of the sensor array to be tightly coupled to data collection at high sampling rates and real-time computation. For high resolution imaging, ultrasonic measurements involve sensors operating in the MHz frequency range. This results in a significant computational load for embedded systems. In order to address these challenges, we present an exploration of architectural variations and realizations that are designed for the efficient achievement of real-time performance goals and constraints. A smart sensor system is presented for ultrasonic pulse-echo measurements targeting flaw detection and parametric echo classification applications. For real-time implementation, an FPGA-based system is developed on Xilinx Virtex-5 FPGA platform. In order to achieve high throughput and robust performance, several design methodologies are investigated; these include the execution of the flaw detection and echo classification algorithms on an embedded microprocessor configured within the FPGA running a Linux operating system, the creation of a dedicated hardware solution, use of hardware/software (HW/SW) co-design methods, and network enabled operations.

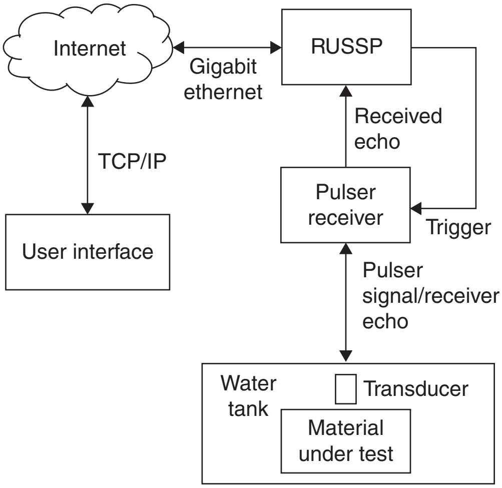
Advances in digital signal processing (DSP) hardware facilitate the development of low-cost and portable ultrasound sensor devices. Instead of requiring multiple discrete components, an ultrasound measurement system can now be built using just DSP hardware and an analog-to-digital converter (ADC). The processing requirements for ultrasonic DSP hardware have traditionally been met using application specific integrated circuits (ASICs) which provide very high performance but have substantial costs and cannot be reconfigured for different applications, thereby limiting their availability to low-volume applications and research. Many ultrasonic applications are now able to fit within the capabilities of reconfigurable logic such as FPGAs (Rodriguez-Andina *et al.*, 2007). An FPGA provides all the parallelism capabilities of an ASIC while allowing the device to be reprogrammed for a given application. This allows a single FPGA to be used in testing multiple algorithms that would traditionally require multiple manufacturing runs (or additional die area) to implement in an ASIC. DSP and FPGAs have opened additional avenues for ultrasound signal analysis techniques in both industry and academia. FPGAs satisfy the demands of researchers due to their relatively low cost, reconfigurability and ASIC-like parallelism and performance capabilities. Within the industry, time to market and budget limitations have forced even shorter development periods. FPGA-based systems are well-suited for smart sensor ultrasonic imaging applications due to having a much shorter development time and lower costs than those of an ASIC.

The proposed RUSSP provides real-time signal analysis and image processing for a full range of ultrasonic testing and imaging applications. The project aims to streamline the development and implementation of signal processing algorithms in embedded software and reconfigurable hardware. This provides the user with an opportunity to explore the full design space including software only, hardware only and HW/SW co-design. Specifically, the RUSSP provides high-speed access to a 12-bit 250 MSPS Maxim MAX1215N ADC (Maxim1215N, 2012) controlled by a Xilinx XC5VLX110T FPGA (Xilinx, 2010). Access to the ultrasound data and custom Internet protocol (IP) cores is available through a gigabit Ethernet connection managed by an embedded Linux-based operating system running on a Microblaze processor instantiated in the FPGA fabric. The major feature of this system-on-chip (SoC) is the integration of a processor which is capable of executing platform independent C code. The ability to execute C code allows the reuse of a substantial portion of available software, thus reducing design and development time. This system also promotes additional reuse at the hardware level as IP cores are able to integrate easily with the processor. Furthermore, the bus architecture has been integrated into an off-the-shelf operating system so that device drivers can be easily reused.

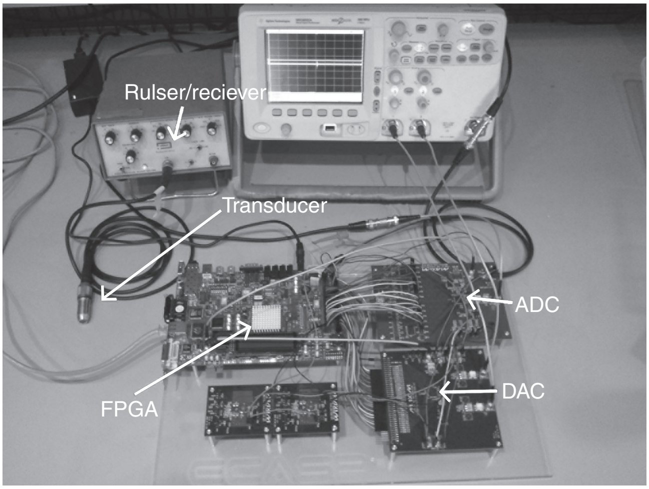
8.2 Fundamentals of ultrasonic sensing and pulse-echo measurements

When an ultrasonic wavelet passes through materials, it loses energy and its sound pressure decreases. This loss results from two causes: scattering and absorption, both of which can be combined in the concept of attenuation. Scattering results when the material is not strictly homogeneous and contains boundaries at which the acoustic impedance changes abruptly due the difference in density and/or sound velocity at the interfaces. There are also materials which are inhomogeneous by their nature. Typically, the degree of inhomogeneity can be estimated in two separate modes of signal analysis. A practical method is by an estimation of the attenuation coefficient. The other widely used method is by inspection of the backscattered echoes. One of the most popular ultrasonic imaging techniques involve transmitting an extremely short burst of sound energy into the sample and monitoring the return echo. This method is referred to as the ‘ultrasonic pulse-echo technique’ for NDE and imaging applications. Depending on the experimental setup and the type of signal processing, one may use these received signals to represent the data in one, two, or three dimensions.

In ultrasonic pulse-echo scanning, the received signal contains the reflected echo from physical and geometrical variation and discontinuity within the sample under testing. The time of flight of these echoes provides information which pertains to the depth of the reflector, and the intensity yields information about the gradient of the impedance and the size of the reflector. In practice, identification of the reflecting surfaces becomes extremely difficult, if not impossible, due the existence of multiple reflections. The setup for ultrasonic pulse-echo measurement using the proposed RUSSP is shown in Figs 8.1 and 8.2.


Internet
Gigabit ethernet
RUSSP
TCP/IP
Received echo
Pulser receiver
Trigger
User interface
Pulser signal/receiver echo
Water tank
Transducer
Material under test

8.1 Typical ultrasound experimental setup using RUSSP for real-time signal processing and imaging applications.


Rulser/reciever
Transducer
ADC
DAC
FPGA

8.2 RUSSP system including Virtex-5 FPGA, ADC, DAC, pulser/receiver and ultrasonic transducer.

The RUSSP is designed to be an easy-to-use platform for developing high throughput, real-time signal processing algorithms. The primary requirement is the use of an ultrasonic pulser to excite the transmitting transducer. This transducer converts the electrical signal to an acoustical radiation field. The generated ultrasonic wave both propagates and reflects echoes corresponding to the characteristic of the propagation path in the sample under test. For example, in NDE of materials there may be cracks or cavities created by stress during use, impurities during manufacturing and so on. Because such cracks and cavities introduce a discontinuity within materials, they can be observed using ultrasonic pulse-echo measurement techniques. The reflected ultrasonic wave excites the receiving transducer which converts the transient vibration energy into an electrical signal. In RUSSP, the backscattered and reflected echoes can be sampled with a precision of 12 bits and up to 250 MSPS. Through experimentation, it was determined that the pulsar/receiver can operate at a maximum trigger rate of 10 kHz.

8.3 Reconfigurable ultrasonic smart sensor platform (RUSSP) design

This section describes the design metrics, system components and the architecture for the RUSSP.

8.3.1 System features and user interface

It is important to build a system that is convenient for both hardware and software developers by meeting the interface requirements. The data acquisition system needs to be integrated with the Matlab programming environment due to common use of the Matlab tool among algorithm developers. Furthermore, Matlab can be used in the validation of the systems hardware and software components. The application programming interface (API) to communicate with the system is based on transmission control protocol (TCP) and IP using the UNIX socket API. The UNIX socket API is a commonly-used and well-understood methodology for transmitting and receiving data across the Internet. The Xilinx Embedded Development Kit software (Xilinx, 2012) is used in developing RUSSP hardware. This allows the use of self-contained *pcores* for hardware components of the system and Xilinx Board Definition (XBD) files to designate the SoC design, off-chip connections and constraints. The self-contained nature of these methods allows reuse across different implementations of the RUSSP. Finally, the RUSSP uses a Linux-based operating system which is accessible over the secure shell (SSH) and serial port (Corbet *et al.*, 2005; POSIX, 2008). This provides a common UNIX shell interface to the developer for system exploration and debugging.

8.3.2 System response and real-time operational requirements

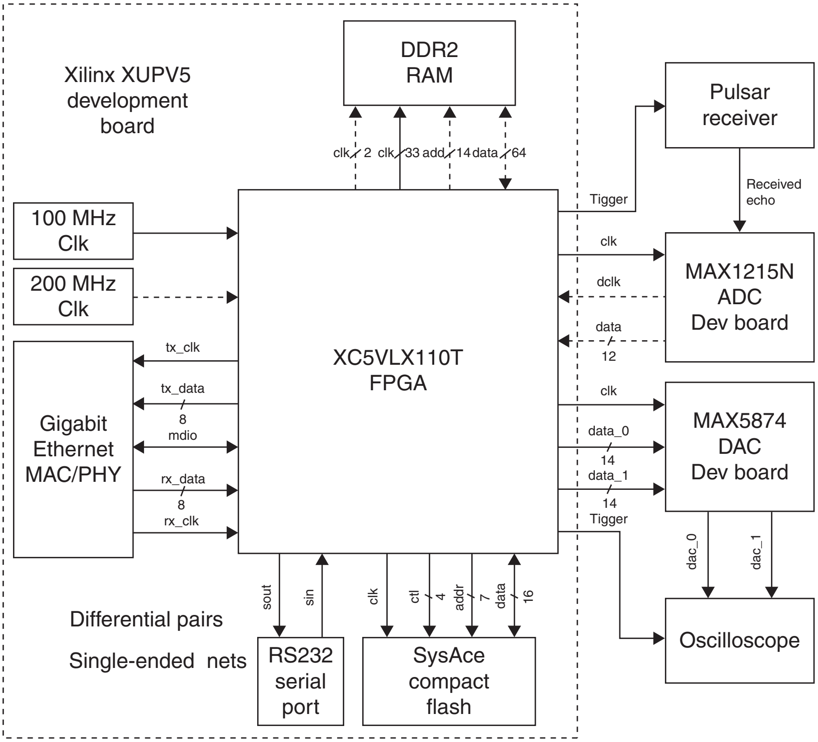
Real-time processing in the DSP context is considered as a system which is capable of processing the incoming signal before the next signal arrives. However, in relation to software and hardware, real-time constraints are: determinism, jitter and delay. Determinism is the requirement of a system to finish processing a required task within a finite time frame. Jitter is the difference between the minimum and maximum response time to an external event. Delay is the maximum time between an external event occurring and a process executing. Both DSP and software/hardware real-time constraints are applicable to the RUSSP system.

The system must provide an interface which offers less than 5 ns jitter for use in coherent averaging since the maximum sampling rate is 200 MSPS. This interface should provide access to the incoming sample, a high-speed memory interface and the time in samples since the pulse was triggered. As the bulk of the system is implemented in FPGA logic, the system is able to guarantee strict hard real-time demands in hardware; however, software interrupt and scheduling latencies should be contained to avoid buffer overruns when copying data from the ADC to the memory or network. The maximum delay for a software application running on the RUSSP is 100 μs or 10 kHz. This ensures that the software is able to process a response before the next response is acquired.

8.3.3 RUSSP architecture

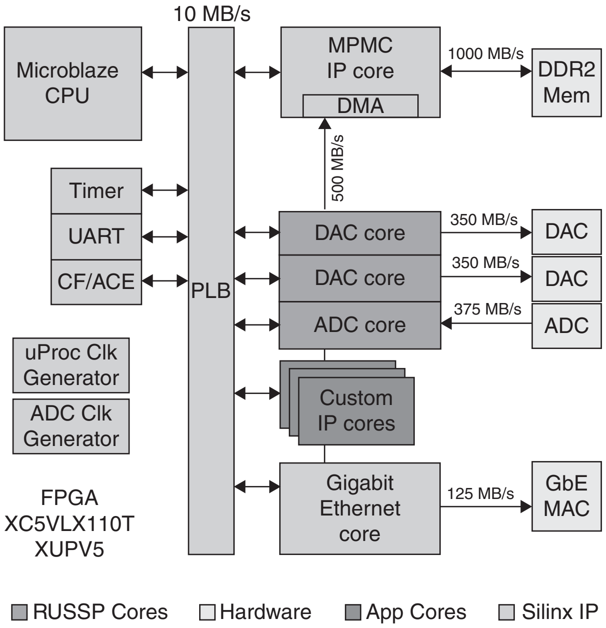
The RUSSP was built using several components including a Xilinx XUPV5 FPGA development platform (Xilinx, 2010), a Maxim MAX1215N ADC Evaluation Kit (Maxim1215N, 2012), a Maxim MAX5874 DAC Evaluation Kit (Maxim5874, 2012) and two Maxim MAX1536 Power Supply Evaluation Kits (Maxim1536, 2012). Integration of these components constructs the RUSSP system. The system can be miniaturized by creating a custom-printed circuit board which includes chips from each of the components above. The interconnections of these components are presented in Fig. 8.3 and the throughput of the various interfaces is presented in Fig. 8.4.

In the RUSSP design, the XUPV5's XC5VLX110T FPGA is used for signal processing. The DDR2 memory controller is used for volatile data storage during processing. The compact flash is used for nonvolatile storage of the FPGA and software configuration. The serial port is used for debugging and configuration. The gigabit Ethernet controller provides remote data and control communication. The benefit of using a Virtex-5 FPGA such as the XC5VLX110T over other FPGAs is the inclusion of DSP48 blocks.


Xilinx XUPV5 development board
DDR2 RAM
clk 2
clk 33
add 14
data 64
Pulsar receiver
Tigger
Received echo
100 MHz Clk
200 MHz Clk
clk
dclk
data 12
MAX1215N ADC Dev board
tx_clk
tx_data 8
mdio
rx_data 8
rx_clk
Gigabit Ethernet MAC/PHY
XC5VLX110T FPGA
clk
data_0 14
data_1 14
Tigger
MAX5874 DAC Dev board
dac_0
dac_1
sout
sin
clk
ctl 4
addr 7
data 16
Differential pairs
Single-ended nets
RS232 serial port
SysAce compact flash
Oscilloscope

8.3 RUSSP block diagram showing hardware components and interconnections.

DSP48 blocks provide constant multiply and accumulate (MAC), two operand multiply and division capabilities. The ability to execute MAC operations greatly increases the maximum operating frequency of digital filter implementations. The system is capable of storing up to 256 MB of data in the DDR2 memory for future analysis. The DDR2-400 used in RUSSP has a maximum theoretical bandwidth of 3200 MB/s and interfaces to the RUSSP though the multi-port memory controller (MPMC) provided by Xilinx (Xilinx, 2011). The gigabit Ethernet controller is connected to the processor using the processor local bus (PLB) interface and a soft direct memory access (SDMA) personality interface modules (PIMs) of the MPMC. The SDMA PIM provides high-speed scatter-gather direct memory access (DMA) transfers to the Ethernet MAC so that the processor is free to process even when data is being copied into or out of main memory. The gigabit Ethernet is connected to the local network and provides ample bandwidth for simultaneous command/control and data capture. The main bottleneck in the RUSSP is the throughput of the PLB bus inside the FPGA SoC.


10 MB/s
Microblaze CPU
MPMC IP core
DMA
1000 MB/s
DDR2 Mem
500 MB/s
Timer
UART
CF/ACE
PLB
DAC core
DAC core
ADC core
350 MB/s
350 MB/s
375 MB/s
DAC
DAC
ADC
uProc Clk Generator
ADC Clk Generator
Custom IP cores
Gigabit Ethernet core
125 MB/s
GbE MAC
FPGA
XC5VLX110T
XUPV5
▨ RUSSP Cores
☐ Hardware
▨ App Cores
☐ Silinx IP

8.4 RUSSP embedded system and bus interfaces showing the maximum bandwidth information. 10 MB/s between CPU and FPGA fabric; 1 GB/s for DDR2 accesses, 350 MB/s for DAC, 375 MB/s for ADC and 125 MB/s for gigabit Ethernet.

Therefore, in future iterations of this design integration of DMA peripherals will be used to reduce data transfers across the PLB bus.

The RUSSP system communicates with the user using TCP/IP on top of gigabit Ethernet. The RUSSP system is theoretically capable of providing wired gigabit Ethernet communication at 125 MB/s. Given a packet of 1500 bytes, the total overhead for IP and Ethernet is approximately 4%. Given a further 1% of overhead for TCP, this leaves a total of 118.75 MB/s for data transmission.

A block diagram of the FPGA embedded system is shown in Fig. 8.4. The FPGA SoC is based around the PLB. PLB is a standard from IBM which defines how components interconnect with one another in a standard bus topology. The SoC includes the following cores from Xilinx: an MPMC, a universal asynchronous receiver transmitter (UART), Gigabit Ethernet, a System Advanced Configuration Environment (Sys ACE) compact flash controller, timer, clock generators and a Microblaze processor.

The MPMC provides high-speed memory access to the system. It provides this through PIM. PIMs allow the memory to transfer data using common patterns. There are PIM modules for PLB, Advanced Microcontroller Bus

Architecture (AMBA), DMA, Frame Buffer, Cachelink, Locallink and a raw native interface. Three PIMs are used in our design. A Cachelink PIM is used to interface the MPMC with the processor caches. A PLB PIM is used to provide access to memory for the PLB bus peripherals. A DMA PIM is connected to the gigabit Ethernet MAC to provide high-speed data transfers.

The software runs on a Xilinx Microblaze processor. The Microblaze processor is a 32-bit reduced instruction set CPU (RISC) which is unique because it allows the engineer to make modifications to its functionality during synthesis. Notably, the processor can be configured to include a memory management unit (MMU) which is a requirement of the Linux kernel. The Microblaze core can be customized to include a floating-point unit (FPU) and streaming interface to either a fast simplex link or an Advanced eXtensible Interface (AXI) stream interfaces. This allows the core to have additional pseudo instructions added to the instruction set. In order to strike a balance between resource usage and performance, 12-KB data and 4-KB instruction caches were chosen. The cache is backed by the MPMC which interfaces to the DDR2 memory.

A timer/counter, interrupt controller and UART are present in the base system to accommodate the requirements of Linux. The timer counter provides timer interrupts to the software system, the interrupt controller allows multiple interrupts to be received and decoded by the software system, and the UART provides a basic interface to the system for debugging.

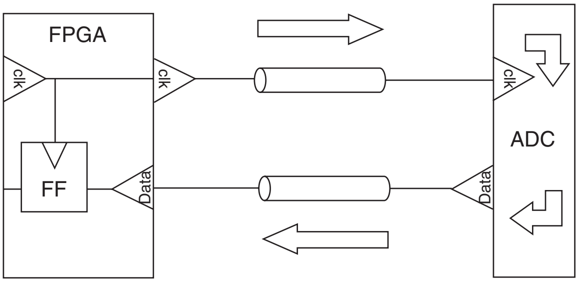
8.3.4 Analog-to-digital converter (ADC) to field programmable gate array (FPGA) interface

In choosing an ADC, the main goal was to design for a broad frequency spectrum of ultrasound from 20 kHz to 20 MHz. The MAX1215N is capable of supporting up to 250 MSPS which is above the minimum required sampling rate. This allows enough bandwidth to capture all the received ultrasonic echoes. Assuming a 250 MHz clock, the total throughput of the ADC is 375 MB/s.

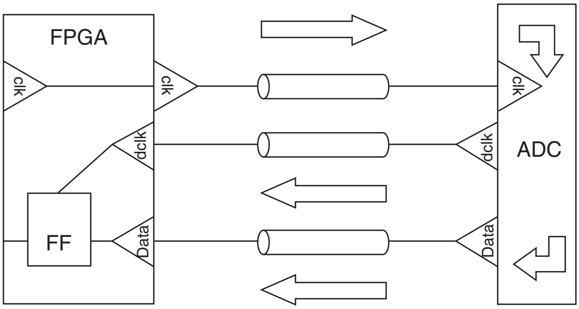
The ADC uses low voltage differential signaling (LVDS) when interfacing to the FPGA. LVDS is a differential signaling methodology which uses a small current to generate a voltage at the receiver. The LVDS generates a 3.5 mA current on either the negative or positive (N or P) net depending on the polarity of the signal. This is then transformed into a voltage using a terminating resistor. The key to making a reliable and high-speed connection between the FPGA and the ADC is to use the DIFF_TERM attribute of the FPGAs input/output block (IOB) for each LVDS pair in the constraints file. This attribute turns the built-in, 100 Ω, differential termination resistor on 100 Ω (Xilinx, 2010).

Initially, the connections between the ADC and the FPGA can be represented in Fig. 8.5. The maximum operating frequency is governed by the delay, t_{skew}, which depends on the length of wire and delay of the ADC. Because the edge rate of the ADC is significant with respect to the length of wire, the line must be treated as a transmission line. The transmission line has a delay of 170 ps/inch according to the measurement obtained in this study which compares an 18-inch *vs* 12-inch coaxial wire. For this arrangement, the total skew is approximately 6.23 ns and limits the connection to a maximum sampling rate of 160.50 MHz.

In order to meet the 200 MHz requirement of the RUSSP system, the recovered clock feature of the ADC was implemented as shown in Fig. 8.6. The recovered clock is aligned with the data which ensures that minimal clock skew is present (assuming equal length data/clock traces). This allows the ADC to operate at its maximum operational frequency.



8.5 ADC to FPGA interconnection without recovered clock. This setup can achieve a maximum sampling rate of 160.50 MHz.



8.6 ADC to FPGA interconnection with recovered clock which achieves minimal clock skew. A sampling rate greater than 200 MHz is possible with this setup.

8.4 Algorithms used in evaluation of RUSSP

Throughout this study, several algorithms and techniques have been implemented in the RUSSP. Coherent averaging is used as a method for reducing electrical noise. Split-spectrum processing (SSP) (Bilgutay *et al.*, 1979; Newhouse *et al.*, 1982; Saniie *et al.*, 1988) is used as a method for reducing Rayleigh scattering noise. Chirplet signal decomposition (CSD) (Lu *et al.*, 2006) is used as a method for classifying and compressing ultrasound data. Combining these methods creates an effective ultrasound signal classification and detection system.

8.4.1 Coherent averaging

Coherent averaging is a method for reducing the signal-to-noise ratio (SNR) of a pulsed signal. It is based on acquiring multiple responses and averaging them together based on their time difference from the start of the pulse. Given an ideal signal, $s(t)$, corrupted by the additive noise, $v(t)$, the received signal, $r(t)$, can be represented as:

$$r(t) = s(t) + v(t) \tag{8.1}$$

The signal is assumed to be deterministic and the same for multiple measurements. The additive noise is a random process with zero mean and noise embedded within multiple measurements that are assumed to be independent and identically distributed. Then, the SNR of the received signal is defined as:

$$\mathrm{SNR} = \frac{\langle S^2(t)\rangle}{[E[v]^2(t)]} \tag{8.2}$$

where $[E[v]^2(t)]$ indicates the expected noise power (i.e. noise variance) and $\langle S^2(t)\rangle$ represents the signal power. Since the target is stationary and the excitation pulse is constantly repeated, the coherent average of multiple responses can be used. Therefore, the SNR of a coherently averaged signal can be expressed by:

$$\mathrm{SNR}_{\mathrm{AVG}} = \frac{\langle S^2(t)\rangle}{\left[\frac{1}{N}E[v]^2(t)\right]} \tag{8.3}$$

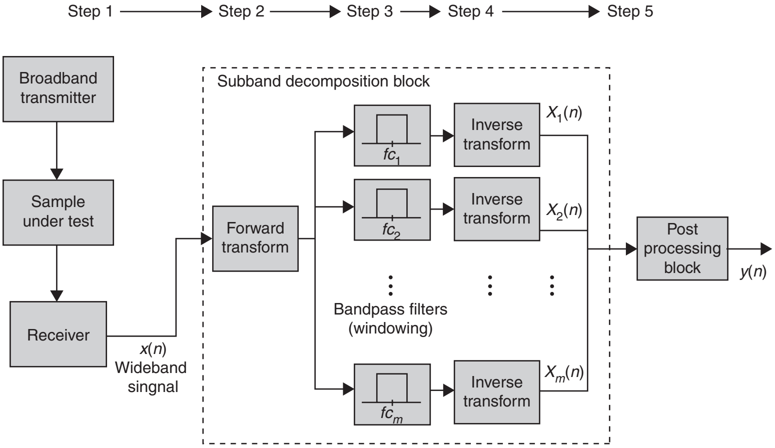
Inspection of $\mathrm{SNR}_{\mathrm{AVG}}$ reveals that a $1/\sqrt{N}$ reduction in noise amplitude is produced by coherent averaging.

8.4.2 Split-spectrum processing

SSP is an algorithm used for discerning frequency-diverse echoes. The algorithm splits a broadband signal into multiple narrowband signals (Bilgutay *et al.*, 1979; Newhouse *et al.*, 1982). These signals are then post processed using algorithms such as absolute minimization, median filtering, or more complex algorithms based on fuzzy set logic, ordered statistics and neural networks. The general goal is to remove the Rayleigh scattering in a system thereby increasing the target-to-clutter ratio (TCR). In a given signal, there are three types of scattering: (i) Rayleigh scattering, (ii) stochastic scattering and (iii) geometric scattering. The relationship between the size of the discontinuity and the wavelength of the pulse determines the type of scattering created. Rayleigh scattering occurs when the object causing the scattering is small compared with the wavelength of the ultrasound pulse. Stochastic scattering occurs when the wavelength of the ultrasound pulse is similar in size to the measured object. Diffusion scattering occurs when the wavelength of the ultrasound pulse is much smaller than the size of the object. Of the three types of scattering, Rayleigh scattering and stochastic scattering can be considered noise for an ultrasonic flaw detection system.

The combined effects of Rayleigh scattering and stochastic scattering are known as 'clutter'. Clutter is a combination of the many echoes that are produced in the microstructures of a material (i.e. grains). When an ultrasound wave scatters from grains, it produces a stationary scatter which represents these grain scatterers. When all of the scatterer echoes are combined, they produce a static noise that hides the target echo. Therefore, these scatterings must be filtered from the signal in order to increase the SNR for the classification or detection system and enable accurate detection of the target. When a measurement is taken using a number of different narrowband pulses with different frequencies, the clutter echoes become randomly distributed across the pulse frequencies and can be filtered using statistical processing. For the performance evaluation of the RUSSP, there is an emphasis on a post processing methodology called 'pass-through absolute minimization'. Pass-through absolute minimization takes advantage of the special properties of absolute minimization (Newhouse *et al.*, 1982) in improving the SNR but maintains the sign of the original signal such that the response can be classified by other signal processing methods, such as CSD.

The SSP procedure has five steps, as shown in Fig. 8.7: (1) data acquisition; (2) fast Fourier transform (FFT) gives the frequency spectrum of the


Step 1
Step 2
Step 3
Step 4
Step 5
Broadband transmitter
Sample under test
Receiver
x(n)
Wideband singnal
Subband decomposition block
Forward transform
fc1
fc2
fcm
Bandpass filters (windowing)
Inverse transform
Inverse transform
Inverse transform
X1(n)
X2(n)
Xm(n)
Post processing block
y(n)

8.7 Ultrasonic SSP algorithm. Step 1 is ultrasonic data acquisition, Step 2 is Fourier transform; Step 3 is sub-band decomposition; Step 4 is inverse Fourier transform; Step 5 is post processing operation – such as minimization.

received echo signal; (3) several filters split the spectrum into different frequency bands; (4) inverse FFT gives the time domain signal of each individual frequency band; (5) the signals from each individual frequency band are first normalized and then passed into a post processing block for detection. This detection processor can employ different techniques, such as averaging, minimization, order statistic filters (e.g. minimization), or Bayesian classifiers (Saniie *et al.*, 1988, 1991, 1992; Saniie and Nagle, 1992). SSP has been successfully applied to ultrasonic flaw/target detection applications due to its robust performance.

8.4.3 Chirplet signal decomposition

Ultrasonic signals are often composed of many interfering echoes. Each echo is similar to a chirplet and chirplets can be described with six parameters using CSD (Lu *et al.*, 2006, 2008). Therefore, CSD representation results in a major data reduction (compression) when compared with the raw ADC data, and can be used for data storage and assessment. Furthermore, the estimated chirplets can be used for material characterization and system identification.

The chirplet signal can be defined as:

$$f_\Theta(t) = \beta \exp\left[-\alpha_1(t-\tau)^2 + i2\pi f_c(t-\tau) + i\Phi + i\alpha_2(t-\tau)^2\right] \qquad [8.4]$$

This equation has six parameters $\Theta = \left[\tau, f_c, \beta, \alpha_2, \phi, \alpha_1\right]$ where term τ is the time-of-arrival, f_c is the center frequency, β is the amplitude, α_2 is the chirp rate, φ is the phase and α_1 is the bandwidth factor of the echo. This model accurately models the echoes from ultrasound experiments. The algorithm for CSD is as follows (Lu *et al.*, 2008):

1. Find the maximum location of $s(t)$ in the time domain and use it as the initial guess of time-of-arrival and the starting point of iteration.
2. Estimate the center frequency which maximizes the chirplet transform, given the initial guess of time-of-arrival.
3. Estimate the time-of-arrival which maximizes the chirplet transform, given the estimated center frequency from the previous step.
4. Estimate the center frequency which maximizes the chirplet transform, given the new estimated time-of-arrival from step 3.
5. Check convergence:
   If $\delta\tau < \tau_{\text{lim}}$ and $\delta f < f_{\text{lim}}$ and $\delta f < f_{\text{lim}}$ where τ_{lim} and f_{lim} are predefined convergence conditions, then go to the next step; otherwise, go back to step 3.
6. Estimate the amplitude β and the remaining parameters α_2, ϕ and α_1 successively.
7. Obtain the residual signal by subtracting the estimated echo from the signal.
8. Calculate the energy of the residual signal (E_r) and check convergence (E_m is a predefined convergence condition): If $E_r < E_m$, STOP; otherwise go to step 1.

The method for extracting chirplet parameters utilizes the chirplet transform:

$$\mathrm{CT}(\Theta) = \int_{-\infty}^{+\infty} f_{\Theta}(t)\, \psi^{*}_{\Theta}(t)\, dt \qquad [8.5]$$

where $\psi^{*}_{\Theta}(t)$ is the chirplet kernel and can be represented as:

$$\psi^{*}_{\Theta}(t) = \left(\frac{2\gamma_1}{\pi}\right)^{\frac{1}{4}} \exp\left(-\gamma_1 (t-b)^2 - i\omega_0\left(\frac{t-b}{a}\right) - i\theta - i\gamma_2 (t-b)^2\right) \qquad [8.6]$$

$\hat{\Theta} = \left[b, \dfrac{\omega_c}{2\pi a}, \eta, \gamma_2, \theta, \gamma_1\right]$ denotes the parameter vector of the chirplet used for transformation.

The first step of this method is to extract an estimated b based on the maximum value in the time domain. Then, the parameter a, which has the highest correlation at $t = b$, can be used as an initial estimate. From these two values, the γ_2 parameter can be estimated, followed by γ_1, θ and β. This equation for the estimation algorithm is summarized below:

$$\left.\begin{array}{c}\dfrac{\partial\,|\mathrm{CT}(\Theta)|}{\partial a} = 0 \\ \hline \dfrac{\partial\,|\mathrm{CT}(\Theta)|}{\partial b} = 0\end{array}\right\} \rightarrow b = \tau \quad \text{and} \quad \frac{\omega_0}{a} = \omega c \qquad [8.7]$$

$$\left.\frac{\partial\,|\mathrm{CT}(\Theta)|}{\partial \gamma_2}\right|_{b = \tau, \frac{\omega_0}{a}} = 0 \rightarrow \gamma_2 = \alpha \qquad [8.8]$$

$$\left.\frac{\partial\,|\mathrm{CT}(\Theta)|}{\partial \gamma_1}\right|_{b = \tau, \frac{\omega_0}{a}, \gamma_2 = \alpha_2} = 0 \rightarrow \gamma_1 = \alpha \qquad [8.9]$$

$$\left.\frac{\partial\,|\mathrm{CT}(\Theta)|}{\partial \phi}\right|_{b = \tau, \frac{\omega_0}{a}, \gamma_2 = \alpha_2, \gamma_1 = \alpha_1} = 0 \rightarrow \phi = \theta \qquad [8.10]$$

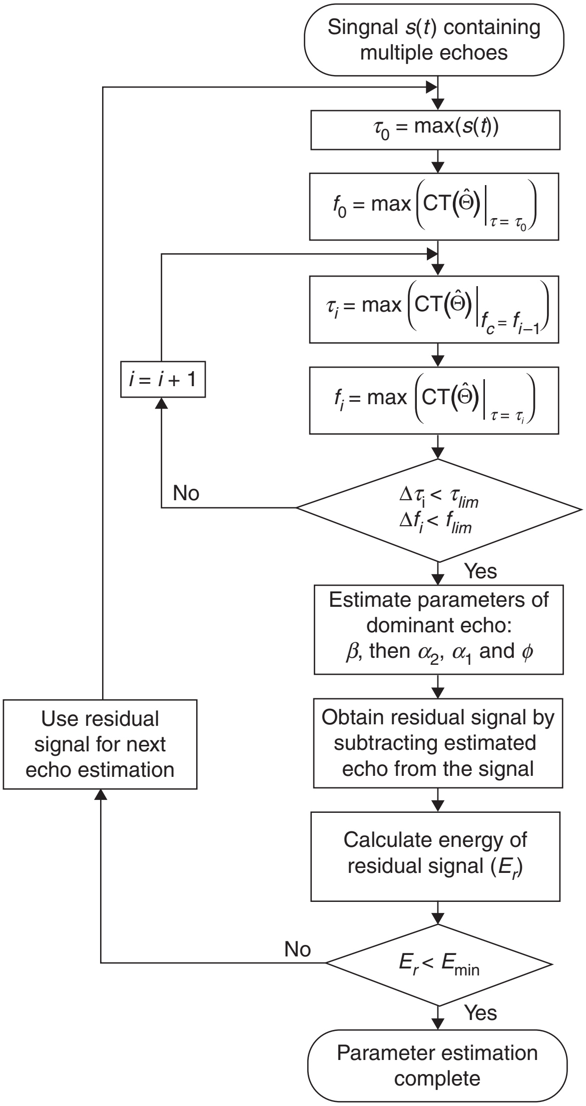
The CSD block diagram for the estimation of chirplet parameters is shown in Fig. 8.8.

8.5 Hardware realization of ultrasonic imaging algorithms using RUSSP

RUSSP is a flexible system designed to host real-time ultrasonic signal processing algorithms. Due to the reconfigurable logic, an embedded CPU and the Linux OS, RUSSP facilitates multiple implementation schemes such as HW/SW co-design for fast design time and rapid prototyping. This section presents implementation details for four case studies including the algorithms discussed in Section 8.4.

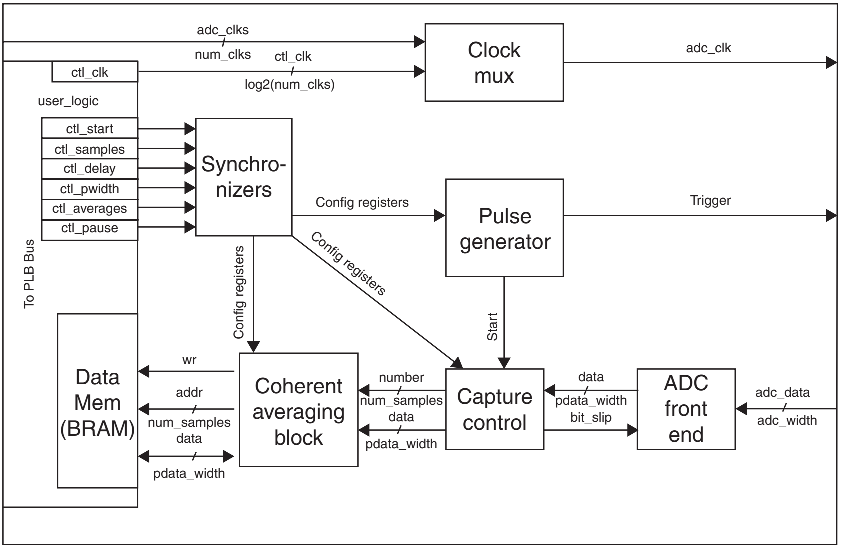
8.5.1 Averaging implementation

Coherent averaging relies heavily on precise synchronization between the pulse and capture logic in order to provide accurate results. Thus, this


Singnal $s(t)$ containing multiple echoes
$\tau_0 = \max(s(t))$
$f_0 = \max\left(\mathrm{CT}(\hat{\Theta})\Big|_{\tau=\tau_0}\right)$
$\tau_i = \max\left(\mathrm{CT}(\hat{\Theta})\Big|_{f_c=f_{i-1}}\right)$
$f_i = \max\left(\mathrm{CT}(\hat{\Theta})\Big|_{\tau=\tau_i}\right)$
$i = i + 1$
No
$\Delta\tau_i < \tau_{lim}$
$\Delta f_i < f_{lim}$
Yes
Estimate parameters of dominant echo: β, then α_2, α_1 and ϕ
Use residual signal for next echo estimation
Obtain residual signal by subtracting estimated echo from the signal
Calculate energy of residual signal (E_r)
No
$E_r < E_{\min}$
Yes
Parameter estimation complete

8.8 CSD flowchart for estimation of ultrasonic echoes. This is based on chirplet model parameters; time-of-arrival τ, center frequency f_c, amplitude β, chirp rate α_2, phase ϕ and the bandwidth factor of the echo α_1.

implementation takes advantage of the preprocessing block in the ADC (as shown in Fig. 8.9) to create a highly accurate synchronization between the data acquisition capture clock and the excitation trigger pulse applied to the ultrasonic transducer. Since the preprocessing block is clocked by the


adc_clks
num_clks
ctl_clk
ctl_clk
log2(num_clks)
Clock mux
adc_clk
user_logic
ctl_start
ctl_samples
ctl_delay
ctl_pwidth
ctl_averages
ctl_pause
Synchro-nizers
Config registers
Pulse generator
Trigger
Config registers
Config registers
To PLB Bus
Start
Data Mem (BRAM)
wr
addr
num_samples
data
pdata_width
Coherent averaging block
number
num_samples
data
pdata_width
Capture control
data
pdata_width
bit_slip
ADC front end
adc_data
adc_width

8.9 ADC system implementation.

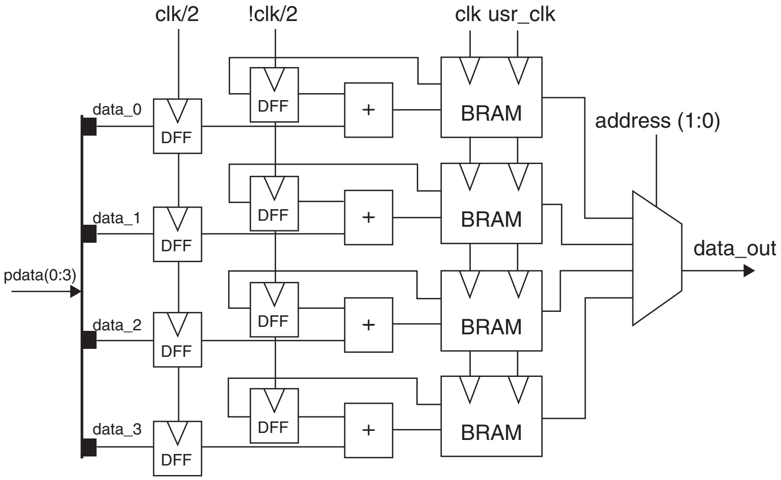
data clock of the ADC, the timing is guaranteed to be accurate to within the frequency of the clock used to drive the ADC.

Initially, a basic averaging implementation was developed. However, it was not possible to meet timing with this original design. In order to explore the design space for timing closure, the original implementation was extended to support parallel operations. In the current configuration, the coherent averaging block is parallelized by a factor of 4 and can be represented by Fig. 8.10. Due to the concurrent nature of this design, the implementation is able reliably to meet the timing requirement in this system.

Since the slower distributed RAM was unable to meet the 200 MHz timing requirement, the coherent averaging block implements its memory using Block RAMs (BRAM.) In Xilinx FPGAs, BRAMs are faster but have a one-cycle delay for access. Therefore, it takes two clock cycles for an accumulation to occur. The first cycle is a read cycle and the second is a write cycle. During the read cycle, the BRAM exports the value to a register. During the write cycle, the incoming data is added to the value in the register and stored back in the BRAM.

8.5.2 Split spectrum processing (SSP) implementation

The goal of SSP is to reduce interference resulting from scatterers. An implementation of the SSP algorithm has been instantiated inside the FPGA fabric. The SSP hardware uses FFT followed by a basic one-zero windowing



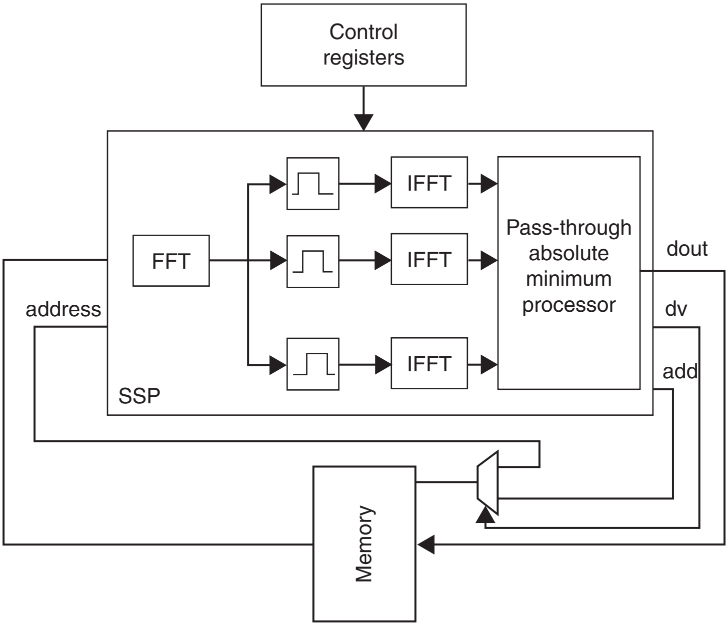
8.10 ADC coherent averaging block. DFF is a D flip-flop and BRAM is the block RAM available in Xilinx FPGA.

algorithm to split spectrum, inverse FFT (iFFT) to obtain frequency-diverse signals, and pass-through absolute minimum processing to improve the visibility of the target echo in the presence of clutter.

As the SSP algorithm uses a significant number of FFTs, it is important that the FFT implementation is highly area efficient. The most efficient IP core available for implementing an FFT/iFFT is the Xilinx Radix 2-Lite FFT IP Core (FFT, 2012). At a transform size of 4096 points, this IP core has a latency of 57 539 cycles; however, it uses only 2 DSP48 components, 7 BRAMs, and approximately 250 slices. The low resource usage allows up to 13 FFTs to be implemented in tandem with the other components of the RUSSP.

The SSP implementation has synthesis parameters which can be altered in the embedded development kit (EDK) for changing the number of channels, and the maximum number of samples. In the current configuration, the SSP algorithm uses 12 channels and 4096 samples, which is the maximum number possible given the DSP48 resources available on the FPGA. For the SSP processing, the number of FFTs required is equal to the number of channels plus 1. In the current configuration, the implementation has 13 FFTs instantiated in the FPGA.

Control registers specify the number of samples to use; and the start, width and offset of the SSP algorithm. The implementation makes use of a memory which is used to load the samples into the SSP algorithm and store them after processing. The memory has two ports; one is connected to the SSP algorithm block, as shown in Fig. 8.11; the other is connected to the PLB bus and is addressable by the processor.



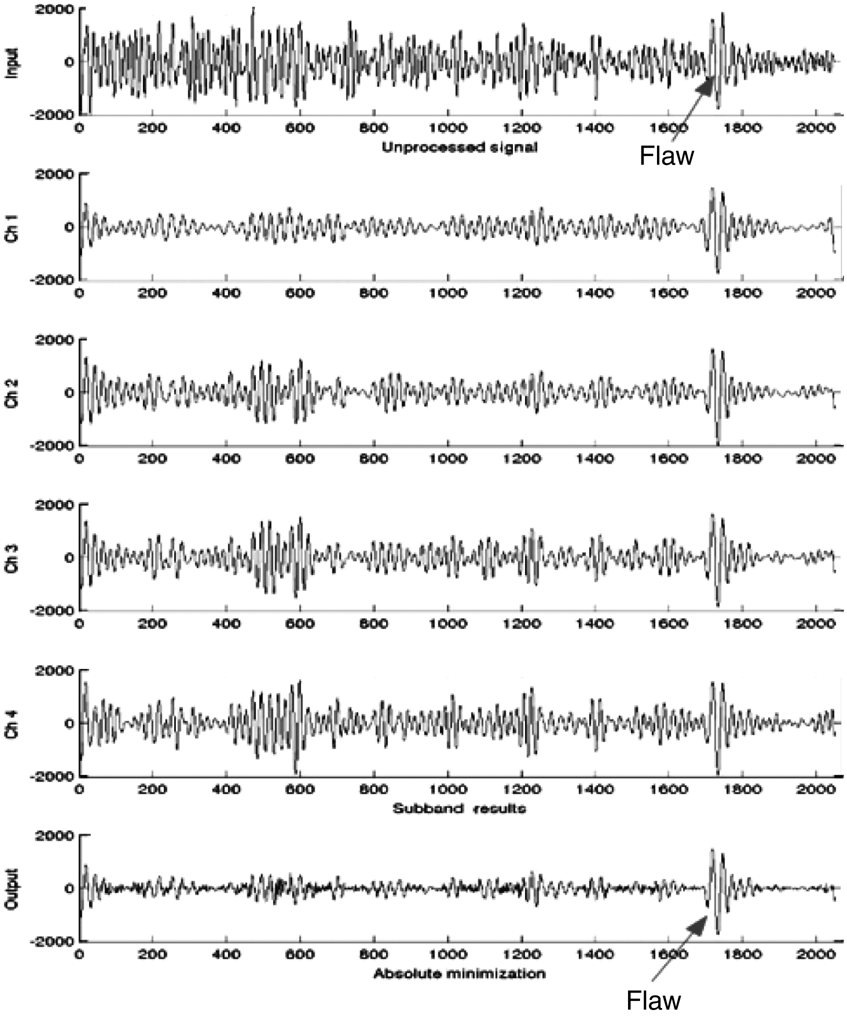
8.11 Block diagram of the SSP implementation.

In Fig. 8.12, the results from a four-channel instantiation of the SSP algorithm are shown. The flaw echo, located at around sample number 1750, is present in each of the channels and has unchanged amplitude. This is because the flaw echo is not a composite signal in contrast to the clutter signal which is highly dependent on the frequency span of narrow-bands. By using a pass-through absolute minimization processing block, the flaw signal maintains many of its original properties, but the frequency dependent scattering signal is substantially reduced.

8.5.3 Chirplet signal decomposition (CSD) implementation

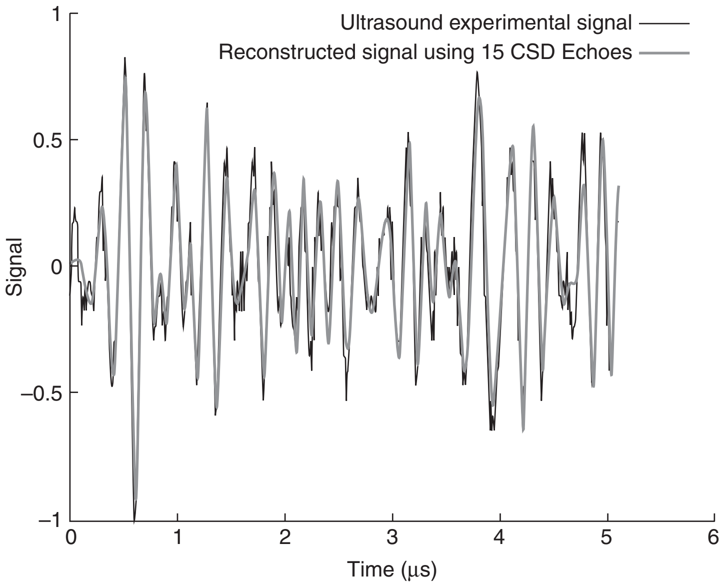
An implementation of the CSD algorithm was created in C programming language. The algorithm was implemented in software due to the high complexity. The basis for the implementation is the algorithm discussed in Section 8.4.3; however, the algorithm has been optimized by using precomputation and estimation methods. Additionally, the algorithm has been altered to provide deterministic behavior.

The primary goal of the current implementation of the CSD algorithm is to reduce the execution time as much as possible, while maintaining an adequate SNR for echo estimation. The main performance impediment for



8.12 SSP sub-bands and results. Channel 1 covers the frequency band 0–4 MHz, Channel 2 covers the frequency band 0.5–5.9 MHz, Channel 3 covers the frequency band 1–5.8 MHz and Channel 4 covers the frequency band 1.5–5.9 MHz.

CSD is the correlation operations which require regenerating a chirplet with different parameters thousands of times during each parameter estimation step in the algorithm. By pre-computing a table of these values and manipulating this table to perform the operation, the computation time was drastically reduced. This was accomplished by finding expressions which could have shared parts and extracting them from the inner loops. Furthermore, this CSD implementation takes advantage of estimation techniques to



8.13 CSD echo extraction and parameter estimation. Top trace: original signal; middle trace: estimated echoes; and bottom trace: reconstructed signal. Vertical axis is normalized amplitude and horizontal axis is time in microseconds.

reduce the overall computation time. By using lookup tables for cosine, sine and tangent, the total execution time was substantially reduced with only a slight reduction in accuracy.

The CSD algorithm is enhanced by altering the nondeterministic operations such as steps 5 and 8 from the listing in Section 8.4.3. In steps 5 and 8, the reiteration of the algorithm if the error condition is not met makes it impossible to determine the total execution time reliably. To gain determinism, the reiteration condition is changed such that a finite number of iterations is performed. In our case, the number of time-of-arrival and frequency re-estimates (step 5) is two and a total of 15 echoes are extracted (step 8). These parameters ensure a total SNR of the regenerated signal equal to approximately 10 dB. In Fig. 8.13, the echoes from the CSD algorithm are regenerated and compared with the echoes in the original signal.

8.5.4 Resource usage and timing constraints

RUSSP implementation uses a significant portion of the FPGA resources when the SSP algorithm is included. The total resource usage of the RUSSP with a 4-, 8- and 12-channel SSP implementation is shown in Table 8.1. In the

Table 8.1 RUSSP resource usage with SSP hardware

Resource	Available	Base	4-Ch	8-Ch	12-Ch
DSP48	64	6	23	41	59
SLICES	17 280	6743	8612	10 279	11 326
BRAMS	148	43	66	81	97

Table 8.2 CSD execution times and SNR for a dataset of 512 samples

Echoes	Execution time (ms)	Signal-to-noise (dB)
1	1050	1.365
2	2050	2.073
4	4040	3.008
8	8060	5.604
15	14880	9.979
16	15940	10.371

case of a 12-channel SSP implementation, a total of 13 hardware FFTs are implemented in the FPGA fabric and a total of 59 DSP48s are employed. The total number of channels for the SSP implementation is limited by the total number of DSP48s available in the system. Removal of the SSP component cuts the resource usage in half.

The RUSSP meets timing for all critical nets in the system. The system has a total of three main clock domains: the processor clock, the ADC clock and the SSP clock. The processor meets timing at 100 MHz, the ADC clock functions at 200 MHz and the SSP clock functions at 100 MHz. The total execution time of the SSP implementation with a total of 2048 samples as measured from the processor is 16 ms. The testing methodology for the SSP execution times is based on the clock_gettime(CLOCK_MONOTONIC) function. This function returns wall clock time in a monotonic fashion such that alterations due to network time protocol or other clock tuning would have no effect on the results.

Execution times for the CSD algorithm are listed in Table 8.2. These results show a nearly linear increase in execution time with respect to the number of echoes.

8.6 Future trends

For embedded ultrasonic signal processing systems, new reconfigurable hardware platforms – such as the Xilinx Zynq-7000 Extensible Processing Platform (EPP) (Zynq-7000 2012) – present an important evolutionary step. Zynq combines an industry-standard ARM® dual-core Cortex™-A9

MPCore™ processing system with Xilinx 28 nm programmable logic. This platform would be ideal for RUSSP applications since the embedded ARM processor provides a significant boost to the software execution capabilities. Additionally, due to the penetration of embedded ARM devices, it is more convenient to acquire code development tools, technical support and market acceptance.

Dynamic partial reconfiguration (DPR) also provides an interesting approach to developing applications for the RUSSP. DPR allows users to download hardware accelerators into the system during run-time (Rossi *et al*., 2009, 2010; Sudarsanam *et al*., 2010; Ahmed *et al*., 2011; Desmouliers *et al*., 2011). This allows applications to swap their accelerators in and out as FPGA resources and requirements allow. In a sense, DPR achieves time-multiplexing of the available reconfigurable fabric for multi-tasking operations. Several ultrasonic platforms already support DPR (Yoon *et al*., 2006; Desmouliers *et al*., 2008; Oruklu and Saniie, 2009), and it is expected to gain wider adoption in the near future with more mature tools and support from FPGA vendors.

8.7 Conclusion

The RUSSP is an adaptable FPGA-based platform for implementing real-time ultrasonic signal processing applications. The architecture of the RUSSP makes it adaptable effectively to implement new algorithms, analyze old algorithms and optimize a system for embedded system usage. By providing the ability to implement algorithms using hardware, software or co-design approaches, end users are given a large design space to find the optimal implementation of a given algorithm with respect to processing speed, logic area usage and power efficiency. Future iterations of the RUSSP will pave the way for more advanced algorithms and additional avenues for design space exploration of ultrasonic smart sensors.

8.8 Sources of further information and advice

More information about ultrasonic imaging systems can be found in the journal publications; *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, *Journal of the Acoustical Society of America*, *Ultrasonics* (Elsevier) and in the proceedings of the IEEE Ultrasonics Symposium, International Congress on Ultrasonics and Acoustical Society of America meetings.

8.9 References

Ahmed, W., Shafique, M., Bauer, L. and Henkel, J. (2011). mRTS: Run-time system for reconfigurable processors with multi-grained instruction-set extensions.

In *Design, Automation Test in Europe Conference Exhibition*, Grenoble, France, 14–18 March, pp. 1–6.

Bilgutay, N., Saniie, J., Newhouse, V. and Furgason, E. (1979). Flaw-to-grain echo enhancement. In *Proceedings of IEEE Ultrasound International Conference*, 26–28 September, pp. 152–157.

Corbet, J., Rubini, A. and Kroah-Hartman, G. (2005). *Linux Device Drivers*, Beijing and Sebastopol, CA: O'Reilly Media.

Desmouliers, C., Oruklu, E. and Saniie, J. (2008). Universal wavelet kernel implementation using reconfigurable hardware. In *ReConFig'08 IEEE International Conference on Reconfigurable Computing and FPGAs*, 3–5 December, pp. 373–378.

Desmouliers, C., Oruklu, E. and Saniie, J. (2011). Discrete wavelet transform realisation using run-time reconfiguration of field programmable gate array (FPGA)s. *IET Circuits, Devices & Systems*, **5**(4), 321–328.

Fast Fourier Transform (FFT) IP core (2012). Xilinx datasheet, http://www.xilinx.com/products/intellectual-property/FFT.htm.

Goldsmith, A.M., Pedersen, P.C. and Szabo, T.L. (2008). An inertial-optical tracking system for portable, quantitative, 3d ultrasound. In *IEEE Ultrasonics Symposium*, 2–5 November, pp. 45–49.

Kunita, M., Sudo, M. and Mochizuki, T. (2008). Range measurement using ultrasound fmcw signals. In *IEEE Ultrasonics Symposium*, 2–5 November, pp. 1366–1369.

Lu, Y., Demirli, R., Cardoso, G. and Saniie, J. (2006). A successive parameter estimation algorithm for chirplet signal decomposition. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, **53**(11), 2121–2131.

Lu, Y., Oruklu, E. and Saniie, J. (2008). Fast chirplet transform with FPGA-based implementation. *IEEE Signal Processing Letters*, **15**, 577–580.

MAX1215N (2012). Maxim Datasheet, http://datasheets.maxim-ic.com/en/ds/MAX1215N.pdf.

MAX1536 (2012). Maxim Datasheet, http://datasheets.maxim-ic.com/en/ds/MAX1536.pdf.

MAX5874 (2012). Maxim Datasheet, http://pdfserv.maxim-ic.com/en/ds/MAX5874.pdf.

Newhouse, V., Bilgutay, N., Saniie, J. and Furgason, E. (1982). Flaw-to-grain echo enhancement by split-spectrum processing. *Ultrasonics*, **20**, 59–68.

Oruklu, E. and Saniie, J. (2009). Dynamically reconfigurable architecture design for ultrasonic imaging. *IEEE Transactions on Instrumentation and Measurement*, **58**(8), 2856–2866.

POSIX, Portable Operating System Interface Base Specifications (2008). IEEE Standard for Information Technology – IEEE Std 1003.1–2008.

Rodriguez-Andina, J.J., Moure, M.J. and Valdes, M.D. (2007). Features, design tools, and application domains of fpgas. *IEEE Transactions on Industrial Electronics*, **54**(4), 1810–1823.

Rossi, D., Campi, F., Deledda, A., Spolzino, S. and Pucillo, S. (2009). A heterogeneous digital signal processor implementation for dynamically reconfigurable computing. In *IEEE Custom Integrated Circuits Conference*, *CICC '09*, San Jose, CA, 13–16 September, pp. 641–644.

Rossi, D., Campi, F., Spolzino, S., Pucillo, S. and Guerrieri, R. (2010). A heterogeneous digital signal processor for dynamically reconfigurable computing. *IEEE Journal of Solid-State Circuits*, **45**(8), 1615–1626.

Saniie, J. (1992). Performance evaluation of frequency-diverse Bayesian ultrasonic flaw detection. *Journal of the Acoustical Society of America*, **91**, 2034.

Saniie, J., Donohue, K.D., Nagle, D.T. and Bilgutay, N.M. (1988). Frequency diversity ultrasonic flaw detection using order statistic filters. In *Proceedings of IEEE Ultrasonics Symposium,* vol. 2, pp. 879–884.

Saniie, J. and Nagle, D.T. (1992). Analysis of order-statistic CFAR threshold estimators for improved ultrasonic flaw detection. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, **39**(5), 618–630.

Saniie, J., Nagle, D.T. and Donohue, K.D. (1991). Analysis of order statistic filters applied to ultrasonic flaw detection using split-spectrum processing. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, **38**(2), 133–140.

Sudarsanam, A., Barnes, R., Carver, J., Kallam, R. and Dasu, A. (2010). Dynamically reconfigurable systolic array accelerators: A case study with extended Kalman filter and discrete wavelet transform algorithms. *Computers Digital Techniques, IET*, **4**(2), 126–142.

Weber, J., Oruklu, E. and Saniie, J. (2011). FPGA-based configurable frequency-diverse ultrasonic target detection system. *IEEE Transactions on Industrial Electronics*, **58**(3), 871–879.

Xilinx (2010). Virtex-5 User Guide, http://www.xilinx.com/support/documentation/user_guides/ug190.pdf.

Xilinx (2011). LogiCORE IP Multi-Port Memory Controller (MPMC) v6.03 Datasheet, http://www.xilinx.com/support/documentation/ip_documentation/mpmc.pdf.

Xilinx (2012). Platform Specification Format Reference Manual v13.4, http://www.xilinx.com/support/documentation/sw_manuals/xilinx13_4/psf_rm.pdf.

Yoon, S., Oruklu, E. and Saniie, J. (2006). Dynamically reconfigurable neural network hardware design for ultrasonic target detection. In *IEEE Ultrasonics Symposium*, pp. 1377–1380.

Zynq-7000 (2012). Extensible Processing Platform, Xilinx, http://www.xilinx.com/publications/prod_mktg/zynq7000/Product-Brief.pdf.

9

Advanced optical incremental sensors: encoders and interferometers

S. J. A. G. COSIJNS and M. J. JANSEN,
ASML, The Netherlands

DOI: 10.1533/9780857099297.1.230

Abstract: This chapter discusses the principles of displacement metrology by optical incremental sensors. The chapter first reviews the basic concepts of displacement interferometer systems, their signal processing and error sources. The chapter then discusses the concepts of optical encoder systems and some of their design considerations.

Key words: displacement interferometer, imaging encoder, interferential encoder.

9.1 Introduction

Within the precision engineering and manufacturing industries, displacement laser interferometers and optical encoders are often used as feedback sensors in numerically controlled systems for reliable and accurate non-contact measurement of linear and rotary motion. Laser interferometers use the laser wavelength as a reference. Stabilized Helium-Neon (HeNe) lasers are used in many national standards laboratories and measurement institutes for obtaining a practical traceable reference to the SI unit of length, the metre. Such systems offer an accurate and effective means of delivering traceability for length and dimensional measurement. With the wavelength of light being used as a standard for length metrology, laser displacement interferometers are mostly suited for traceable calibration and acceptance tests of machine tools and coordinate measurement systems. As the wavelength of the laser light is dependent on the refractive index of the air, interferometers are typically used in well-conditioned metrology laboratories. In high-tech industrial equipment, like lithography machines and reticle/mask metrology tools for the semiconductor industry, laser interferometers are often applied for displacement measurement with sub-nm resolution. Optical encoders use an optical grating on a substrate as a reference and are available for both low-end and high-end applications where environmental conditions may be less well-defined, or where there is a risk that a laser beam would be interrupted. In this chapter, the basics of incremental displacement interferometers and optical encoders are explained, together with some of their advantages and disadvantages.

9.2 Displacement interferometers

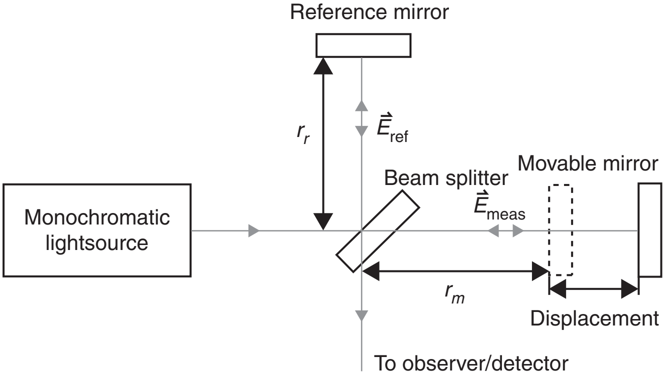
9.2.1 Basics of displacement interferometry

The displacement interferometer, first introduced by Albert Michelson in 1881 (Fig. 9.1), has been developed into a measurement system with high accuracy. Since interferometry is based on the interference of light, it is a non-contact measurement method of which the accuracy is influenced by the wavelength of the source and the medium in which the measurement takes place.

In an interferometer coherent light is directed to a semi-transparent mirror that acts as an amplitude-dividing beam splitter. Part of the light is transmitted towards a movable mirror and reflected by this mirror. The other part of the light is reflected at 90° towards a fixed reference mirror, reflected and recombined at the beam splitter where their interference is observed. The electromagnetic waves propagating in the reference and measurement arm can be represented as:

$$\begin{aligned}\overrightarrow{E}_{\text{ref}} &= E_r \mathrm{e}^{i(\omega t + k\overrightarrow{r_r} - \Phi_r)}\overrightarrow{\mathrm{e}_r}\\ \overrightarrow{E}_{\text{meas}} &= E_m \mathrm{e}^{i(\omega t + k\overrightarrow{r_m} - \Phi_m)}\overrightarrow{\mathrm{e}_m}\end{aligned} \qquad [9.1]$$

with E_r and E_m the amplitude of the respective E-fields, ω the angular frequency, t the time, k the propagation factor $(k = 2\pi/\lambda)$, $\vec{r}$ the position vector,



9.1 Schematic representation of Michelson's interferometer.

ϕ_r and ϕ_m the phase in respectively the reference and measurement arms. After recombination in the beam splitter, the electromagnetic field is the linear superposition of these waves. Assuming the propagation of the field in only one dimension, the irradiance at the detector becomes:

$$I = \epsilon_0\, c\left\langle \vec{E}^2 \right\rangle = \epsilon_0\, c\left(E_r^2 + E_m^2 + 2E_rE_m \cos(k(r_m - r_r) - (\phi_m - \phi_r)\right) \qquad [9.2]$$

where ϵ_0 is the vacuum permittivity, c the speed of light, k is the propagation factor and r_m is the traveled optical distance in the measurement arm and r_r is the traveled optical distance in the reference arm. The constant $\epsilon_0\, c$ is omitted in further calculations for convenience. If the waves were initially in phase $(\phi_r = \phi_m)$, the cosine term depends on the difference in optical path length between the reference and measurement arms. This is the case for a monochromatic light source. If, further, the two beams are of equal amplitude, the irradiance is:

$$I = 2I_0\left(1 + \cos\left(\frac{2\pi}{\lambda}(r_m - r_r)\right)\right) \qquad [9.3]$$

When now the measurement mirror is displaced over a distance Δl while the reference mirror remains fixed, the optical path length changes $2n\Delta l$, with n the refractive index of the medium through which the light travels. The factor 2 is due to the fact that this distance is travelled twice by the light. If a detector is used, the measurement signal will change as follows:

$$I = 2I_0\left(1 + \cos\left(\frac{2\pi}{\lambda}(2n\Delta l)\right)\right) \qquad [9.4]$$

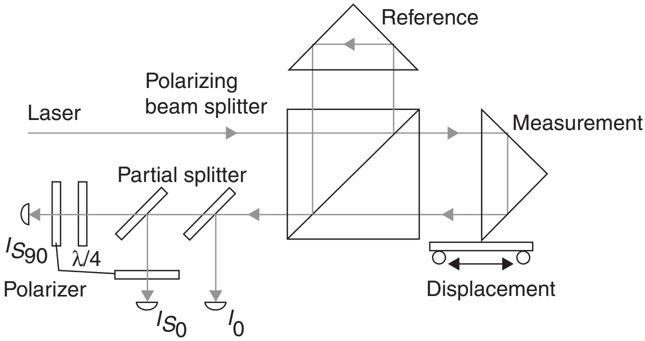
If the wavelength of the light source is known then the displacement can be calculated from the change in intensity on the detector. From this, it can also be seen that it is a relative measurement, only displacement can be measured, not distance.

Numerous different versions were derived from Michelson's original displacement interferometer, all working on the principle of measuring displacement with use of interference. Most modern displacement interferometers use an HeNe-laser as a light source, due to the long coherence length and relatively short visible wavelength which results in comfortable alignment and an improved resolution. Laser interferometers can be divided into two types: homodyne and heterodyne.

Homodyne interferometers (detection)

Most commercial homodyne laser interferometers consist of a stabilized single frequency laser source, polarizing optics, photo detector(s) and measurement electronics. A homodyne laser source is typically a HeNe-laser with a single frequency beam as the output consisting of either a linear polarization under 45° or a circularly polarized beam. The beam is split into the reference arm and measurement arm of the interferometer by a beam splitter. Following reflection from their respective targets, the beams recombine in the beam splitter. In order to observe interference, both beams should have equal polarizations. This is accomplished by a polarizer oriented at 45° to the beam splitter. The photo detector signal is run through electronics which count the fringes of the interference signal. A fringe is a full cycle of light intensity variation, going from light to dark to light. Every fringe corresponds to an optical path difference of a wavelength corresponding to Equation [9.4]. Since there is no intrinsic time dependency in the measurement signal, this is also known as a 'DC interferometer'. Depending on the detector configuration, direction sensing and insensitivity to power changes can be derived as well as a compensation for periodic deviations such as that developed by Heydemann (1981). In Fig. 9.2, the principle of such a homodyne interferometer is shown. Signal I_0 is used to normalize the intensity. Signals s_0 and s_{90} are used for phase quadrature measurement. Signal s_0 is the normal signal of a homodyne interferometer (Equation [9.4]) with signal $2I_0$ subtracted:

$$s_0 = 2I_0 \cos\left(2\pi\left(\frac{2n}{\lambda}\right)\Delta l\right) \tag{9.5}$$



9.2 Schematic representation of the principle of a homodyne laser interferometer with power compensation and direction sensing. I_0 is used to eliminate effects of power changes, S_0 and S_{90} are used for phase quadrature measurement.

In order to enable direction sensing and a constant sensitivity over an entire wavelength, part of the measurement signal is split off and receives a phase shift of 90°; this measurement signal is called s_{90}:

$$s_{90} = 2I_0 \sin\left(2\pi\left(\frac{2n}{\lambda}\right)\Delta l\right) \qquad [9.6]$$

Using these equations, the displacement can be calculated from both intensity signals using:

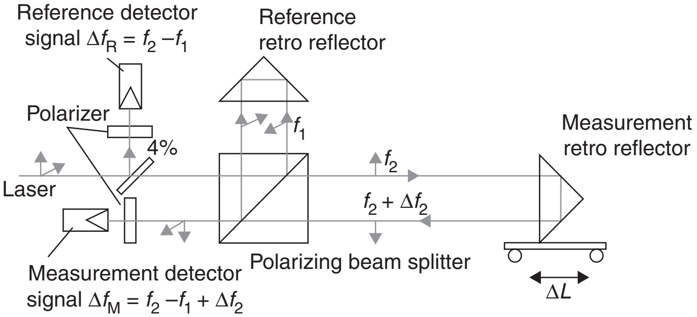
$$\Delta l = \left(\frac{\lambda}{2n}\right)\cdot\frac{1}{2\pi}\tan^{-1}\frac{s_{90}}{s_0} \qquad [9.7]$$

with $(\lambda/2n)$ being the signal pitch, which is the measured displacement corresponding to a phase cycle of 2π rad. In an ideal interferometer, when both signals are plotted against each other a circle is described.

Heterodyne interferometers (detection)

The basic setup of a heterodyne interferometer is shown in Fig. 9.3. Generally the light source of a heterodyne laser interferometer is a stabilized HeNe-laser whose output beam contains two frequency components (f_1 and f_2), each with a unique linear polarization. Their electromagnetic field is represented by:

$$\begin{aligned}\overrightarrow{E_1} &= E_{01}\mathrm{e}^{\mathrm{i}(2\pi f_1 t+\phi_{01})}\overrightarrow{\mathrm{e}_1}\\ \overrightarrow{E_2} &= E_{02}\mathrm{e}^{\mathrm{i}(2\pi f_2 t+\phi_{02})}\overrightarrow{\mathrm{e}_2}\end{aligned} \qquad [9.8]$$


Reference detector signal $\Delta f_R = f_2 - f_1$
Reference retro reflector
Polarizer
4%
f_1
Measurement retro reflector
Laser
f_2
$f_2 + \Delta f_2$
Measurement detector signal $\Delta f_M = f_2 - f_1 + \Delta f_2$
Polarizing beam splitter
ΔL

9.3 Schematic representation of the principle of a heterodyne laser interferometer. The reference retro reflector is in a steady position, while the measurement retro reflector is attached to the object of which the displacement should be measured.

where E_{01} and E_{02} represent the amplitude, and ϕ_{01} and ϕ_{02} represent the initial phase of the electromagnetic field. The frequency difference can be generated by a Zeeman laser. The Zeeman technique produces two frequencies by applying an axial magnetic field to the laser tube. Another way to generate a frequency shift is by an acousto-optic modulator; for example, a Bragg cell driven by a quartz oscillator. With the Zeeman technique, the frequency split is limited to a maximum of ±4 MHz. The acousto-optic modulator enables a frequency split of 20 MHz or more. Generally, in a heterodyne laser interferometer the two polarizations used are orthogonal to each other.

Part of the light emitted by the laser source is split off, passes a combining polarizer and falls onto a detector with a band-pass filter. The resulting signal is an alternating signal with a beat frequency equal to the split frequency in the laser head. This signal forms the reference measurement I_r, given by:

$$I_r = 2E_{01}E_{02}\cos\left(2\pi(f_2 - f_1)t + (\phi_{02} - \phi_{01})\right) \qquad [9.9]$$

As can be seen from Equation [9.9], the heterodyne interferometer works with a carrier frequency (f_2–f_1), therefore it is also known as an 'AC interferometer'. The rest of the light emerges from the laser head and enters the interferometer optics (Fig. 9.3), consisting of a polarizing beam splitter and two retro reflectors. In the polarizing beam splitter, the two frequencies are split by means of polarization. Frequency f_1 is reflected by the polarizing beam splitter and enters the reference arm, then is reflected by the reference retro reflector and is again reflected by the polarizing beam splitter. Frequency f_2 is transmitted by the polarizing beam splitter and enters the measurement arm, then is reflected by the moving retro reflector and is again transmitted by the beam splitter. Ideally, both frequencies emerge from the polarizing beam splitter in their own unique polarization orthogonal to each other. To enable interference, the beams are transmitted through a polarizer at 45° with their polarization axes. After the polarizer, the light falls onto a second detector with a band-pass filter resulting in the measurement signal:

$$I_{\mathrm{m}} = 2E_{01}E_{02}\cos\left(2\pi(f_2 - f_1)t + (\phi_{02} - \phi_{01}) + (\phi_{\mathrm{meas}} - \phi_{\mathrm{ref}})\right) \qquad [9.10]$$

where $\phi_{\mathrm{meas}} - \phi_{\mathrm{ref}}$ is the difference in phase between the signal in the measurement arm and the reference arm. It consists of a constant term in the reference arm (ϕ_{0r}) and a phase in the measurement arm consisting of a constant term (ϕ_{0m}) combined with a changing term as a result of the moving retro reflector:

$$\phi_{\mathrm{meas}} - \phi_{\mathrm{ref}} = \Delta\phi + \phi_{0m} = \phi_{0r} \qquad [9.11]$$

In a practical application, the measurement retro reflector is attached to the object of which the displacement is to be measured. As a retro reflector moves in the measurement arm with velocity v, a Doppler shift is generated for frequency f_2:

$$\Delta f = \frac{2vnf_2}{c} \quad [9.12]$$

with v the velocity of the moving retro reflector, n the refractive index of the medium through which the light travels (e.g. air) and c the speed of light in vacuum. From Equation [9.12], it follows that the maximum traveling speed of the target is limited due to the finite frequency shift between the two frequencies in the laser source. The effective phase change in the interference signal resulting from the Doppler shift equals:

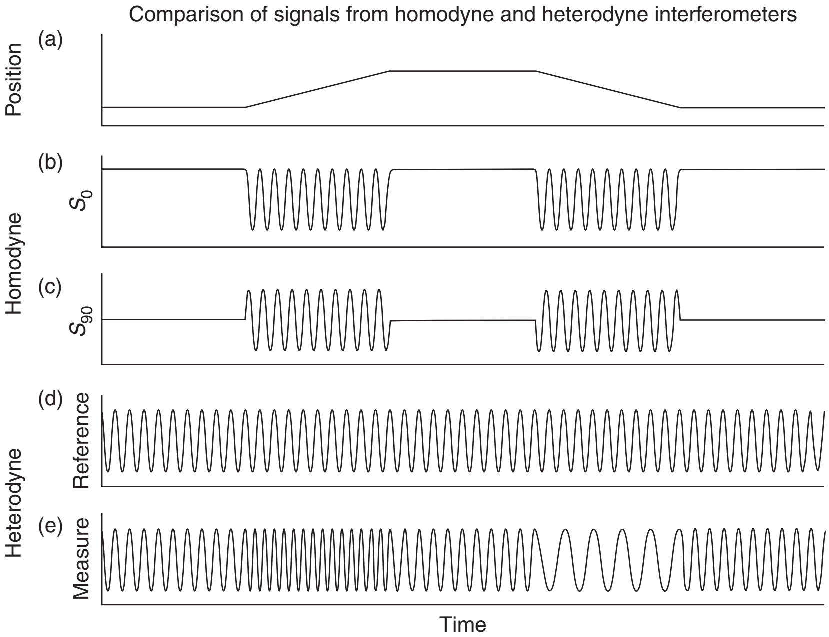
$$\Delta\phi = \int_{t1}^{t2} 2\pi\Delta f dt = \int_{t1}^{t2} 2\pi \frac{2vnf_2}{c} dt = 4\pi \frac{nf_2}{c} \int_{t1}^{t2} v dt = \frac{4\pi nf_2}{c} \Delta l \quad [9.13]$$

where Δl is the displacement of the retro reflector. So, by measuring the phase change between the measurement signal (Equation [9.10]) and the reference signal (Equation [9.9]), the displacement of the retro reflector can be determined by using the inverse of Equation [9.13] with vacuum wavelength λ_2:

$$\Delta l = \frac{\lambda_2}{2n} \cdot \left(\frac{\Delta\phi}{2\pi} \right) \quad [9.14]$$

Signals

Figure 9.4 shows the effect of a moving object on a homodyne and a heterodyne interferometer signal. Comparing the intensity signals for homodyne (Equations [9.5] and [9.6]) and heterodyne interferometers (Equations [9.9] and [9.10]) the main difference lies in the intensity signal during standstill of the measurement object; in the case of a homodyne interferometer, the intensity signal remains constant. In the case of a heterodyne interferometer, the intensity is modulated by the split frequency of the laser, resulting in a carrier frequency. Taking the signals into account shows the principal difference between homodyne and heterodyne detection: During standstill the signal-to-noise ratio is easily disturbed for a homodyne interferometer by any change in the individual signal strength.



9.4 Schematic comparison of the relevant signals of a displacement (a) measurement with a homodyne ((b) and (c)) and a heterodyne interferometer, both the reference signal (d) and the measurement signal (e).

9.2.2 Interferometer concepts

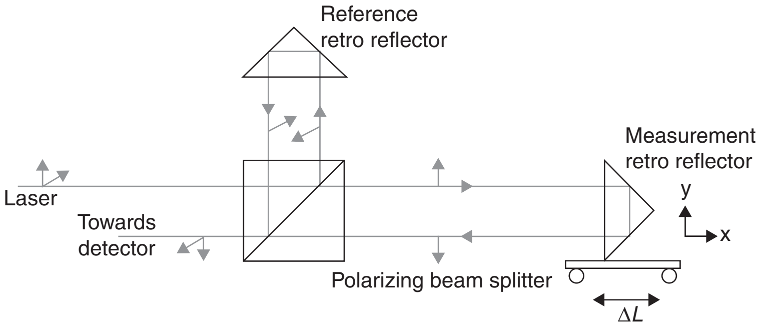
With the basic principles as described in the previous paragraph, different optical configurations can be used to enhance measurement resolution.

Linear interferometer

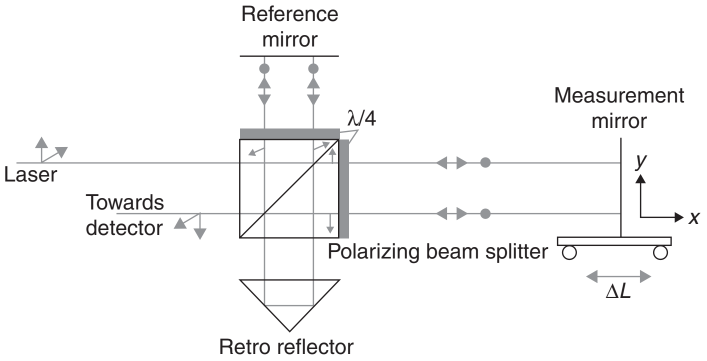
The 1D distance interferometer (see Fig. 9.5) consists of a polarizing beam splitter and two retro reflectors. This type of interferometer is mostly used on a μm scale. The advantage of this sensor is the limited sensitivity for angle variations of the retro reflectors. The disadvantage is the limited range in the direction perpendicular to the measurement direction. As a result this measurement configuration is limited to a 1D movement of the measurement object only.

Plane mirror interferometer

Another commonly used interferometer configuration, known as a 'plane mirror interferometer', is shown in Fig. 9.6. It uses a plane mirror as the


Reference
retro reflector
Measurement
retro reflector
Laser
Towards
detector
y
x
Polarizing beam splitter
ΔL

9.5 Schematic representation of the optics in a linear interferometer.


Reference
mirror
λ/4
Measurement
mirror
Laser
Towards
detector
y
x
Polarizing beam splitter
ΔL
Retro reflector

9.6 Schematic representation of the optics in a plane mirror interferometer.

target, which enables movement of the target in a direction perpendicular to the measurement beam. As a result, this interferometer is often used in configurations where 2D measurements of the moving object are required, such as 2D moving stages. Due to its configuration, the beams travel towards the measurement mirror twice and, as a result, the measurement resolution of the system is doubled.

The light entering from the source is split by the polarizing beam splitter. One polarization is transmitted towards the measurement arm and one is reflected towards the reference arm. The transmitted beam (vertical polarization) passes a quarter-wave plate changing the polarization into circularly polarized light. This light travels towards the measurement mirror where it is reflected and the polarization rotated 180°. The light again passes the quarter-wave plate where it is changed to horizontal polarization and, as a result, now is reflected by the polarizing beam splitter. The light travels to the retro reflector and is returned to the beam splitter, where it is reflected again, passes the quarter-wave plate, reflects a second time on the measurement mirror and passes the quarter-wave plate for the fourth time, resulting in vertical polarized light again and, as a result, is again transmitted by the polarizing beam splitter towards the detector.

The originally horizontal polarization is reflected by the polarizing beam splitter and enters the reference arm, where it is transmitted by the quarter-wave plate resulting in circularly polarized light, which is reflected by the reference mirror and rotated 180°. It then passes again the quarter-wave plate again, resulting in vertically polarized light which is transmitted by the beam splitter toward the retro reflector. There, the light is translated, again enters the beam splitter, passes though the quarter-wave plate for the third time, and is reflected a second time by the reference mirror. Finally, it passes the quarter-wave plate a fourth time and is once more horizontally polarized and reflected by the polarizing beam splitter toward the detector.

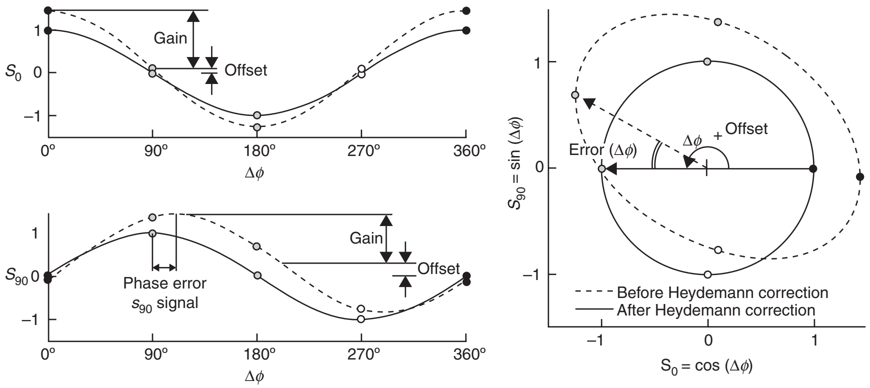
9.2.3 Phase detection and interpolation

Both incremental displacement interferometers and optical encoders output a signal with a phase that varies periodically with displacement Δl.

$$\Delta\phi(\Delta l) = 2\pi\left(\frac{\Delta l}{P}\right) + \alpha \qquad [9.15]$$

where $\Delta\phi$ represents the phase difference between the measurement beam and the reference beam, P the signal period and α a phase offset that is specific for a certain detector signal. Most industrial displacement interferometer and linear encoder systems at least internally generate analog phase quadrature signals consisting of a cosine and sine signal with a 90° phase offset. This phase offset allows for the unambiguous detection of the direction of motion and the sinusoidal nature guarantees a measurement sensitivity that is independent of the measurement position. The resolution of an interferometer system can be increased by phase interpolation. With analog quadrature signals, it is possible to interpolate within each quadrature cycle. The measured phase is represented by the vector angle of the data point on the Lissajous curve (Fig. 9.7).

To reduce periodic errors which repeat every phase cycle, the analog phase quadrature signals are normalized by adjusting the gain, the DC offset and phase offset until a centered circle is obtained when plotting the s_0 and s_{90} signals against each other in a Lissajous plot. In modern systems, the normalization of the Lissajous curve is corrected mathematically by means of a correction like the Heydemann correction (Heydemann,



9.7 Heydemann correction and Lissajous curve. Left: Phase quadrature signals before and after normalization using Heydemann correction to compensate for gain, offset and signal phase errors (Heydemann, 1981). Right: Lissajous curve.

1981). To determine a normalization correction, the measurement signal must, preferably, undergo at least a few phase cycles – typically obtained by introducing a movement of the system in the measurement direction. Correct compensation of DC signal offsets will resolve first-order interpolation errors, whereas proper gain and phase offset calibration corrects for second-order periodic interpolation errors. Depending on the frequency response of a detector, the calibration of the Heydemann correction parameters may have some level of speed dependency, which, if measurement uncertainties in the order of a nanometer are desired, can make optimal correction under all measurement conditions challenging. Phase interpolation is typically obtained by direct calculation of the arctangent function by a microprocessor.

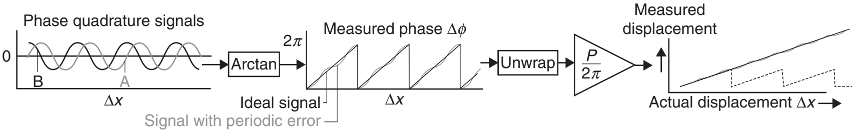
$$\Delta\phi = \tan^{-1}\left(\frac{A}{B}\right) \qquad [9.16]$$

where parameters A and B are the normalized signals that are 90° out of phase and can be derived by taking a linear combination of the interferometer signals in accordance with Table 9.1.

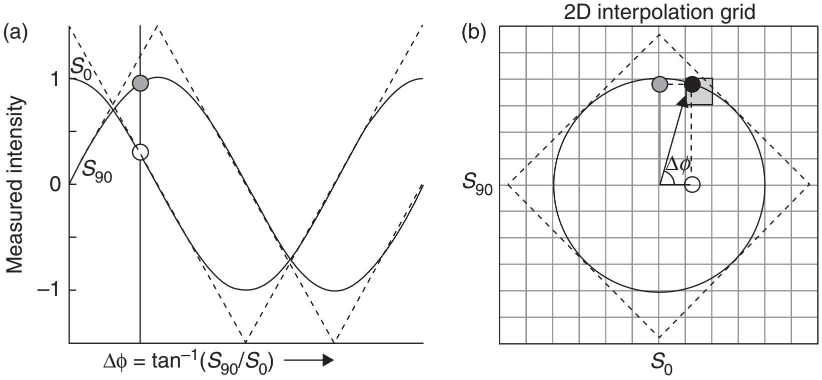
The error sensitivity for laser power and signal amplitude variations may be improved by an optical design that generates three or more analog signals with known phase offsets and of which the average intensity is invariant for laser power fluctuations (i.e. 4 × 90° phase shifts or 3 × 120° phase shifts).

Table 9.1 Determining two quadrature signals *A* and *B* with a 90° phase from a linear combination of sensor signals.

Analog sensor output signals	*A*	*B*
0°, 90° phase offsets	S_{90}	S_0
0°, 90°, 180°, 270° phase offsets	$S_{90} - S_{270}$	$S_0 - S_{180}$
0°, 120°, 240° phase offsets	$\sqrt{3}(S_{120} - S_{240})$	$-S_{240} + S_0 - S_{120}$


Phase quadrature signals
0
B
A
Δx
Arctan
2π
Measured phase Δϕ
Ideal signal
Signal with periodic error
Δx
Unwrap
P/2π
Measured displacement
Actual displacement Δx

9.8 Phase interpolation using direct arctangent calculation and the calculation of displacement.

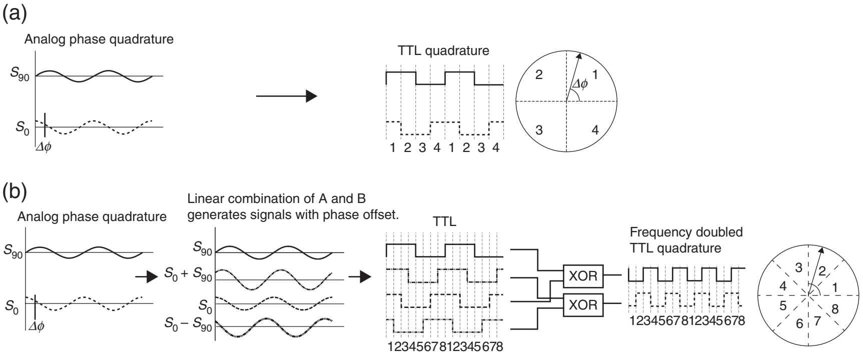


9.9 (a, b) 2D interpolation using a 2D lookup table with pre-calculated arctangent values or binary TTL signal values.

After the interpolated phase is calculated using the arctangent function, the measured displacement can be obtained by unwrapping the measured phase and by multiplication of the unwrapped phase by the signal pitch $P = \lambda/2n$ (see Fig. 9.8).

For time-critical high-speed systems, the interpolated phase may be obtained through a 2D lookup table (Fig. 9.9) (Hagiwara, 1992). This lookup table may be corrected for known signal offsets or nonlinearities which may cause the signal to deviate from an ideal sinusoidal signal.

Some encoders have a digital square wave as an output. This can be derived directly by thresholding the analog sinusoidal signals to obtain a resolution


(a)
Analog phase quadrature
S_{90}
S_0
$\Delta\phi$
TTL quadrature
1 2 3 4 1 2 3 4
2
1
$\Delta\phi$
3
4
(b)
Analog phase quadrature
S_{90}
S_0
$\Delta\phi$
Linear combination of A and B generates signals with phase offset.
S_{90}
$S_0 + S_{90}$
S_0
$S_0 - S_{90}$
TTL
1234567812345678
XOR
XOR
Frequency doubled TTL quadrature
1234567812345678
3
2
4
1
5
8
6
7

9.10 Signal generation examples. (a) TTL signal generation by thresholding the sinusoidal phase quadrature signals resulting in a resolution of one-quarter of the signal pitch.
(b) Phase interpolation using additional phase signals and logic circuits to obtain a frequency doubled TTL quadrature signal with a resolution of one-eighth of the signal period.

of one-quarter of the signal pitch, or the resolution can be enhanced further using interpolating circuits. Additional phase signals with a phase offset α are created by taking a linear combination of the analog cosine (s_0) and sine (s_{90}) signals:

$$\sin(\Delta\phi + \alpha) = \cos(\alpha) s_{90} + \sin(\alpha) \cdot s_0 \qquad [9.17]$$

In the example in Fig. 9.10b, two additional phase signals are generated ($\alpha = 45°$ and $\alpha = 135°$) to obtain a 2× interpolated quadrature signal. Using comparators, the signals with various phase offsets are converted into transistor-transistor logic (TTL) signals. Then, a logic circuit turns these TTL signals into a quadrature signal with increased frequency to obtain a resolution of one-eighth of the signal period. When more additional phase signals are generated, higher levels of interpolation can be obtained by this method (Benzaid and Bird, 1993). In heterodyne detection, the measured phase is transmitted as a phase modulated signal on top of the phase of a reference signal. The phase difference between the beat frequency of the reference signal and the beat frequency of the measurement signals can be detected using a phase-lock-in amplifier, which typically converts the signal with the carrier frequency into a homodyne type s_0 and s_{90} signal. The heterodyne detection method makes the transmitted signal insensitive for DC and gain offset variations of the detector signal for static measurements, but does not eliminate periodic errors as caused by polarization mixing or misalignment of optical components.

9.3 Sources of error and compensation methods

The theoretical resolution of the system is determined by the degree of interpolation of the sinusoidal signals to determine the value $\Delta\phi$, while in practice precision is limited by the quality of the measurement signal. In the preceding description of an ideal interferometer (see 9.2.1), it was shown that the displacement could be determined by Equation [9.7] for a homodyne laser interferometer and by Equation [9.14] for a heterodyne laser interferometer. From these equations, it can be seen that the accuracy of the calculated displacement depends on the accuracy of the determination of the phase change $\Delta\phi$, the wavelength of light used λ and the refractive index of the medium n. Apart from these inherent sources of error, there are numerous other error sources depending on the setup used. The errors in the laser interferometer can be divided into three categories: setup dependent, instrument dependent and environment dependent. In the following paragraphs, a brief explanation is given of these categories and the contributors.

9.3.1 Setup dependent

Depending on the measurement configuration, several errors might occur; for example, cosine errors, Abbe errors, dead path errors, target uniformity and mechanical stability.

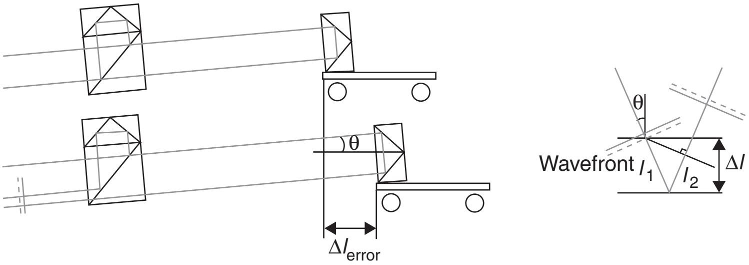
Cosine error

The cosine errors result from an angular misalignment between the interferometers measurement beam and the axis of displacement (see Fig. 9.11). Due to reduced signal efficiency, ultimately a shorter distance is measured than that of the actual displacement; the measurement wavelength is virtually increased. This can be seen on the right-hand side of Fig. 9.11: the detector plane is determined by the wavefront angle of the returning beam. A mechanical path change of Δl_{error} is measured by optical path length $l_1 + l_2$.

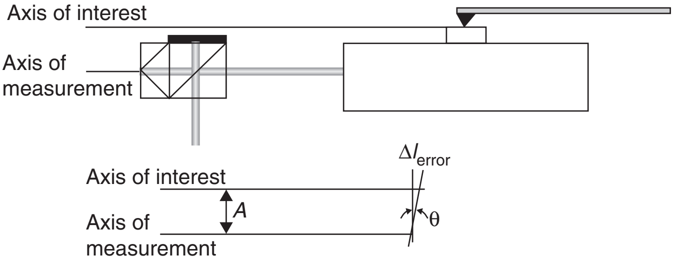
$$\Delta l_{\text{error}} = \Delta l \cos\theta \qquad [9.18]$$

Abbe error

An Abbe error exists if the axis of measurement does not coincide with the axis of movement, as shown in Fig. 9.12. In this case, the movement of an object in the atomic force microscope is to be measured. The interferometer measurement axis does not, however, coincide with the axis of interest


θ
Δl_{error}
θ
Wavefront
l_1
l_2
Δl

9.11 Schematic representation of a cosine error occurring in an interferometer measurement.


Axis of interest
Axis of measurement
Δl_{error}
Axis of interest
A
θ
Axis of measurement

9.12 Schematic representation of an Abbe error in an AFM measurement.

but has an Abbe arm A as the offset, resulting in an error Δl_{error} if the stage is tilted:

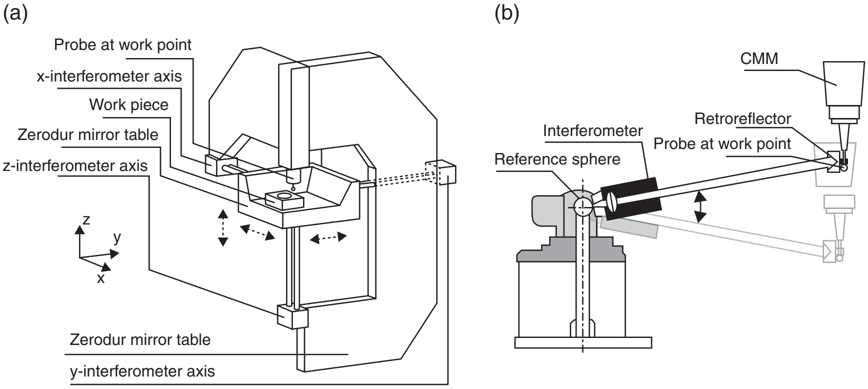
$$\Delta l_{\text{error}} = A\tan(\theta) \tag{9.19}$$

An example of a measurement setup corrected for this Abbe error is shown in Fig. 9.13a. In the interferometer-based nano coordinate measuring machine (CMM) configuration, the measurement beams of three plane mirror interferometers are all pointing towards the work point (stylus tip) to eliminate Abbe errors for *x, y* and *z* displacement measurements (Peggs, 1999; Ruijl, 2001).

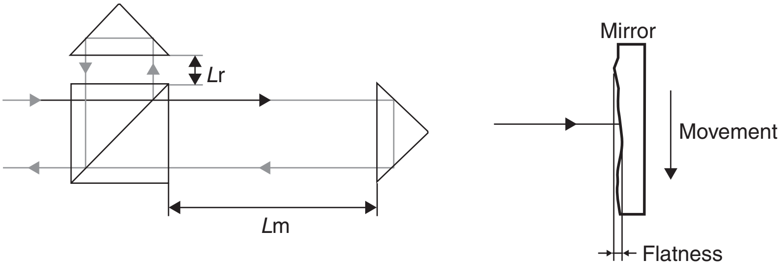
Another example of a laser interferometer that adheres to the Abbe principle is the laser tracker shown in Fig. 9.13b, where the laser beam automatically follows a reflector attached to the probe of a CMM or to the tool head of a production machine (Kniel and Schwenke, 2007).

Dead path error

Dead path errors are errors caused by uncompensated path length of the interferometer as shown in Fig. 9.14, if L_m represents the smallest distance


(a)
Probe at work point
x-interferometer axis
Work piece
Zerodur mirror table
z-interferometer axis
z
y
x
Zerodur mirror table
y-interferometer axis
(b)
CMM
Retroreflector
Interferometer
Probe at work point
Reference sphere

9.13 Ultra precision machines without Abbe errors, (a) using three plane mirror interferometers, all pointing to the work point (Ruijl, 2001). The system has a thermally stable measurement frame of invar and zerodur mirror table. (b) Laser tracker.


Lr
Lm
Mirror
Movement
Flatness

9.14 Schematic representation of dead path error ($L_m - L_r$) (left) and target uniformity (right).

to the measurement target and L_r is the fixed reference arm length, then the dead path is represented by $L_m - L_r$. Due to environmental changes, the optical path length in the dead path changes and creates an error:

$$\Delta l = \Delta n\left(L_m - L_r\right) \qquad [9.20]$$

Target uniformity

If the position of the measurement beam shifts over the measurement target, target uniformity also plays a role. In a situation where a flat mirror is used with flatness $\lambda/20$, without calibration this error can contribute to an error of 32 nm ($\lambda = 633$ nm).

Mechanical stability

As an interferometer will only measure the position of the reflecting target, the stability of the measurement target also contributes to the uncertainty budget of the system. As a result, the stiffness and thermal stability of the target and intermediate body should also be taken into account.

9.3.2 Instrument dependent

Based on the choice of the instrument, several instrument dependent errors add to the uncertainty of the measurement as there are: (split) frequency stability, beam walk off, electronics, data age, data age jitter, periodic deviation and ghost reflection and stray light.

(Split) frequency

In interferometers, the measurement reference is the wavelength of the light which is used, as can be seen in Equations [9.7] and [9.14]. From these equations it becomes clear that, with a frequency change of the interferometer source, a measurement uncertainty is introduced. Usually, the relative stabilities of the laser sources are in the range of 2×10^{-9} and 5×10^{-8}.

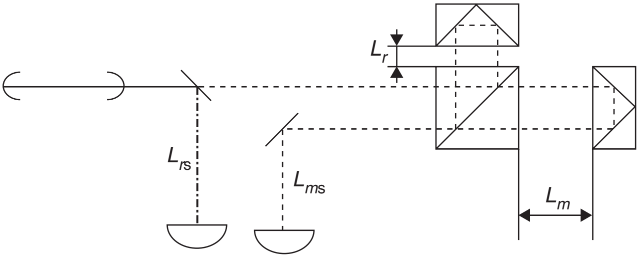
It can be shown that for heterodyne interferometers not only the absolute frequency stability is of importance, but also the stability of the split frequency as shown in the following equation:

$$\Delta l = \frac{1}{n}\frac{(L_{\mathrm{ms}} - L_{\mathrm{rs}})}{p}\frac{f_s}{f_m} + (L_m - L_r)\frac{f_a}{f_m} \quad \text{with} \quad f_a = \frac{f_m + f_r}{2},\ f_s = f_m - f_r \qquad [9.21]$$

where L_{ms} represents the optical path of the entire measurement signal and L_m represents the optical path of the measurement arm as shown in Fig. 9.15, f_m and f_r represent the frequencies nominally in the reference arm and in the measurement arm, and p represents the optical resolution (number of passes) of the interferometer optics.

Beam walk off

Beam walk off is the effect of a rotation of the moving target outside its plane in combination with a different path length of reference and measurement arm (see also Fig. 9.11). As a result, a phase error is generated and the wavefront of the reference arm and measurement arm no longer completely coincide over the full measurement range. Consequentially, the signal-to-noise ratio deteriorates and eventually the measurement range is



9.15 Schematic representation of relevant beam parts for frequency stability effects.

limited by this effect. This especially holds for stage rotation of plane mirror interferometers where the amount of walk off scales with:

$$\mathrm{WO} = 4\phi\Delta l \qquad [9.22]$$

Electronics and data age

Typically the noise and linearity of the electronics will contribute to the uncertainty of the measurement. The actual contribution in uncertainty depends on the optical resolution of the interferometer (number of passes to the measurement target).

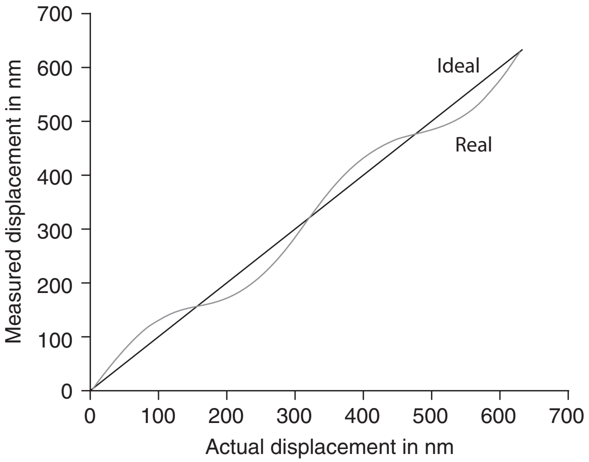
Depending on the velocity (v), the aging of the measurement data (δ) may become a critical part in the measurement error:

$$\Delta l = \delta v \qquad [9.23]$$

For a movement at a speed of 1 m/s, a data delay of 10 ns will generate a direction dependent offset of 10 nm. The variation of the data age (jitter) contributes to the uncertainty of the measurement.

Periodic deviation

As displacement interferometers are often based on polarizing optics, the mixing of polarization will introduce periodic deviations in the measurement as shown in Fig. 9.16. These periodic deviations are also often referred to as nonlinearities. Quenelle (1983) first mentioned the existence of periodic deviations as a result of alignment errors between the laser and the optics, Bobroff (1987) showed the existence of periodic deviations as a result of alignment of the polarizing beam splitter together with the quality of the coating. These errors had a frequency of one cycle per one wavelength



9.16 Schematic representation of measurement results with a periodic deviation. The magnitude is exaggerated for clarity.

optical path change. In the same year, Sutton proved the existence of a periodic deviation with two cycles per wavelength optical path change. First- and second-order periodic deviations were defined at this time. Basically, periodic deviations depend on the polarization quality and birefringence of the optics as well as the antireflection coating. The effects can be reduced in a setup by aligning the optical source to the polarizing optics.

Commercial interferometers may use the compensation methods mentioned in paragraph 9.2.3 to compensate for these periodic deviations as they do show up as harmonics on the measurement signal (Heydemann, 1981; Chu and Ray, 2004). On the other hand, there are also interferometer configurations which minimize the periodic deviations by splitting the reference and measurement paths completely (Wu *et al.*, 1999; Joo *et al.*, 2009, 2010; Kim *et al.*, 2010).

9.3.3 Environment dependent

Thermal influence on the interferometer is relevant in this category, as well as the refractive index of the medium through which the light travels (mostly air).

Thermal effects on the interferometer

Thermal expansion of the interferometer optics or mechanics may cause unwanted effects in the measurements. Depending on the setup, these may introduce drift or pointing effects in the measurement.

Table 9.2 Sensitivity of the refractive index of air.

Parameter	Unit	Effect in *n* Birch and Downs	Effect in *n* Bönsch and Potulski	Typical daily variation (in measurement laboratory)	Effect in n
Temperature	K^{-1}	-9.298×10^{-7}	-9.299×10^{-7}	0.1 °C	-9.3×10^{-8}
Pressure	Pa^{-1}	2.684×10^{-9}	2.683×10^{-9}	20 hPa	5.36×10^{-6}
Humidity	Pa^{-1}	-3.63×10^{-10}	-3.706×10^{-10}	1 hPa	-3.66×10^{-8}
Carbondioxide	ppm^{-1}	0	1.447×10^{-10}	100 ppm	1.45×10^{-8}

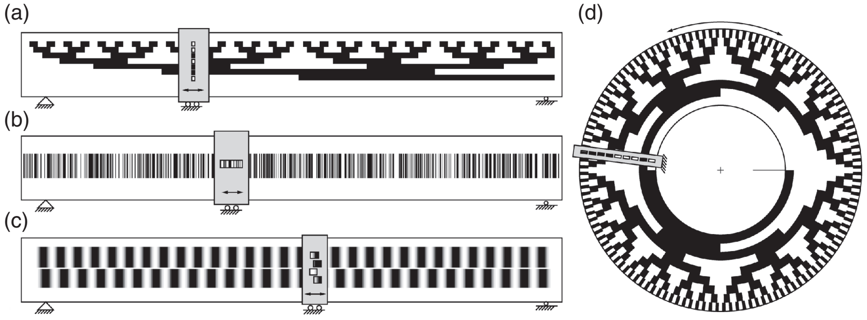
Notes: Depending on a specific parameter according to Birch and Downs (1993), and Bönsch and Potulski (1998) for standard conditions: $T = 20°C$, $p = 101325$ Pa, $f = 926.8$ Pa (40% Rh) and $CO_2 = 450$ ppm.

Refractive index of air

As the wavelength of the system defines the length reference, the stability of the refractive index of the medium through which the measurement occurs determines the accuracy of the measurement. Usually this effect is compensated for using Edlén's updated equation (Edlén, 1966; Birch and Downs, 1993; Bönsch and Potulski, 1998) and measuring temperature, pressure, humidity and the carbon dioxide component. The individual effects are shown in Table 9.2. The relative uncertainty of this compensation method is usually limited to 2×10^{-8}, resulting in a measurement uncertainty of 20 nm/m. Another way to compensate for refractive index effects is by means of a refractometer. In such a system, a reference of vacuum is measured as well as the medium through which the interferometer is measuring in a mechanically stable setup with equal length (Schellekens *et al.*, 1986). In some applications, compensation of the refractive index variations is sufficient, and a wavelength tracker can be used. This tracker usually consists of an extra interferometer measurement with a mechanically stable path length (generally constructed of a low expansion material) which is representative for the nominal measurement length.

9.4 Optical encoders

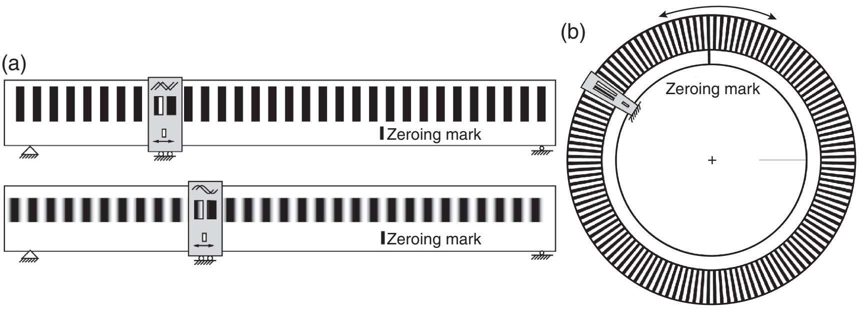
The existing linear encoders encompass various detection techniques, based on brush, magnetic, inductive, capacitive and optical principle. Optical encoders are popular because of their non-contact, high-resolution measurement characteristics and because their output signal is easily converted into an electronic position or displacement signal. The range of applications of optical encoders spans from measurement of angular or linear motion in low-cost consumer products (like computer mice and inkjet printers) to advanced industrial systems (like machine tools, robots and stage



9.17 Schematic representation of absolute linear encoder scales. (a) Gray code, (b) pseudorandom code, (c) parallel readout of subsets of scales with a different pitch, (d) example of an absolute gray-coded optical rotary encoder disk.

positioning in lithographic projection tools). An optical encoder contains a detection system that can detect light emanating from a patterned movable member. The patterned member is usually referred to as the ‘optical scale’. For a linear encoder, the scale typically contains one or more sets of parallel lines with a predetermined constant or varying pitch. Rotary encoders, also known as ‘shaft encoders’, have radial patterns. The detected light is converted into an electrical signal and may be processed as digital square waves or as analog signals, where analog signals allow for resolution enhancement by means of interpolation.

There are two types of optical encoders: absolute position and incremental displacement encoders. Most common absolute position encoders use gray-coded or pseudorandom coded scales or scales with subsets of scales with a different pitch (see Fig. 9.17). The absolute encoders can be used in both linear and rotary configurations. Gray-coded scales have parallel measurement channels for measuring the bits of the gray-coded position information. A pseudorandom coded scale has a unique single track stripe pattern for each absolute position. This unique stripe pattern is read by a line camera and decoded by a microprocessor (Gribble and Robert, 2011). Absolute encoders may be obtained by the combination of phase information of two or more incremental encoders which have a different pitch and are integrated on the same scale. In the example in Fig. 9.17c, the phase differences between the measured phases provide coarse absolute position information which is refined by the interpolated phase of one of the phase quadrature signals. Absolute scales with different pitches are also available as hybrid scales where the coarse absolute position is derived from subsets of capacitance-based scale signals which are then refined by the signal from a fine pitch incremental optical scale (Mitutoyo Catalog, 2012).



9.18 Schematic incremental encoder principles: (a) square/sawtooth output, (b) sinusoidal phase quadrature output, (c) rotary encoder disk.

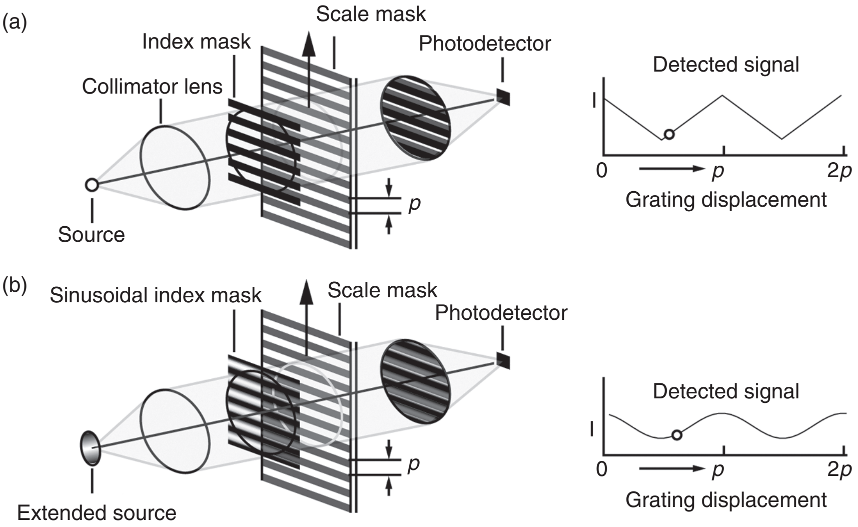
The most commonly used encoders are of the incremental type, which are relatively easy to interface and typically obtain better speed and resolution performance than absolute encoders.

Incremental displacement encoders usually are supplemented with additional reference marks. The measurement of one or multiple uniquely spaced reference marks incorporated in the scale can be used to construct an absolute position signal. If directional information is required, phase quadrature detection is used where at least two (typically 90°) out-of-phase signals are generated. Linear encoders may use transmissive glass scales or reflective scales, employing amplitude gratings or phase gratings to suppress undesired diffraction orders. Scale materials include chrome on glass, metal (stainless steel, gold plated steel, Invar), ceramics (Zerodur, ULE, Clearceram) and plastics. The scale may be self-supporting, mounted via an adhesive film on the backside of the scale, or mounted in a carrier which allows the scale to expand freely at both ends to ensure a defined thermal behavior.

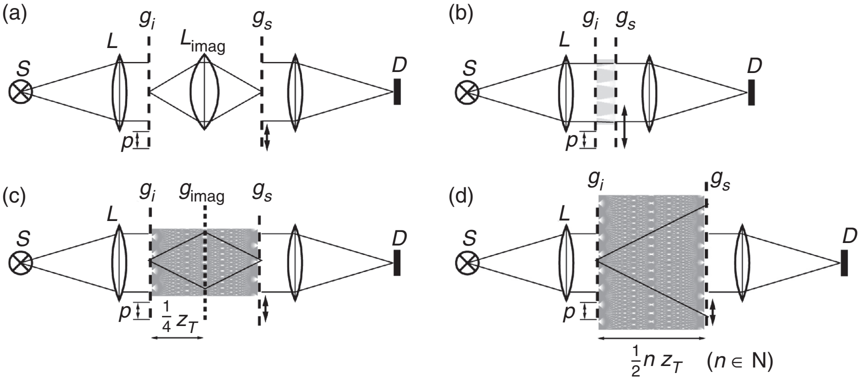
In this chapter, the focus will be on incremental linear encoders as these are the most commonly applied type of optical encoder for high-precision applications. The imaging and interferential grating encoders are the two most common incremental encoder types that create an electrical signal from the relative movement between a scale grating and a reading head.

9.4.1 Imaging incremental encoder

In imaging encoders, light is presumed to travel in a straight line and is typically passed through two periodic patterns forming a moiré pattern. Light is periodically obscured at a photo detector when one pattern is moved relative to the other (see Fig. 9.21). The resolution of optical encoders scales with the grating period. An imaging incremental encoder consists of a light source

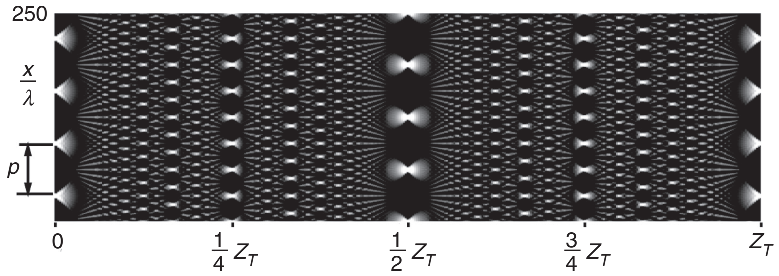

(a)
Scale mask
Index mask
Photodetector
Collimator lens
Source
p
Detected signal
I
0
p
2p
Grating displacement
(b)
Sinusoidal index mask
Scale mask
Photodetector
Extended source
p
Detected signal
I
0
p
2p
Grating displacement

9.19 Imaging moiré encoder: (a) triangular detector signal, (b) sinusoidal detector signal.


(a)
g_i
L_{imag}
g_s
L
S
D
p
(b)
g_i
g_s
L
S
D
p
(c)
g_i
g_{imag}
g_s
L
S
D
p
$\frac{1}{4} z_T$
(d)
g_i
g_s
L
S
D
p
$\frac{1}{2} n z_T$ $(n \in N)$

9.20 Imaging incremental encoder with source *S*, detector *D,* collimator lens *L*. The index grating g_i is imaged onto a scale grating g_s: (a) using an imaging lens L_{imag}, (b) close contact shadow projection, (c) using an imaging grating g_{imag} and adopting the Talbot effect, (d) Fresnel diffraction.

generating a collimated light beam and a stationary index grating g_i which transmits 50% of the light. This transmitted light is imaged onto a moving scale grating g_s either via imaging optics (Fig. 9.20a) or via shadow projection (Fig. 9.20b). Note that the encoder detects relative movement between the index grating and the scale grating. It is therefore not important which grating is identified as the index grating and which the scale grating. If the



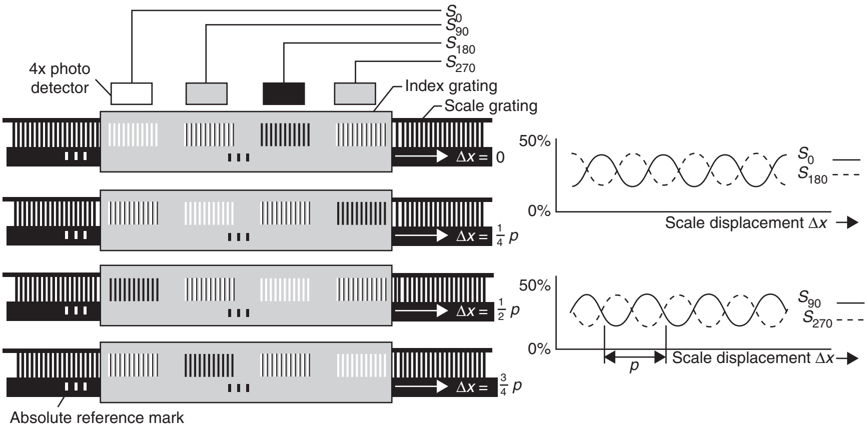
9.21 Talbot effect: self-imaging of a periodic structure that is illuminated by a collimated coherent beam.

index grating is imaged perfectly on top of the transmissive section of the scale grating, then all light is transmitted by the scale grating. If the moving grating is displaced by half a pitch, all light will be obscured by the opaque part of the moving grating. The light that is passed by the scale grating forms a moiré pattern that is detected by a photodetector, which produces an electrical signal that is proportional to the light intensity.

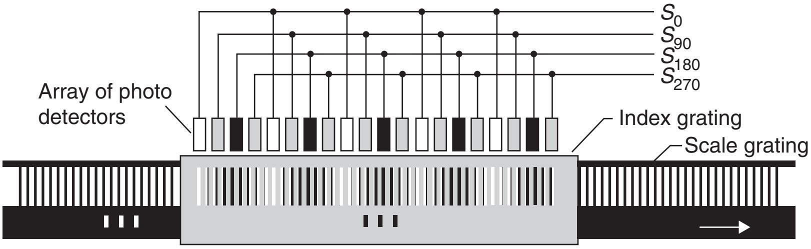
A configuration with an imaging lens allows operation within the depth of field of the imaging optics. A light source, preferably monochromatic, is used to limit the effect of color dispersion. Color dispersion can be avoided when using an imaging grating with a pitch $p/2$ positioned at one-quarter of the Talbot distance z_T (see Fig. 9.21) (Patorski, 1986; Dürschmid, 1993).

The Talbot effect is a near field diffraction effect. Based on Fresnel diffraction, the Talbot effect causes an image of a periodic grating to be reproduced at the Talbot distance $Z_T = 2p^2/\lambda$, where p is the grating pitch and λ is the wavelength of the light incident on the grating. At one-quarter of the Talbot length, the self-image is halved in size, and appears with half the period of the imaged grating. If in Fig. 9.20c the imaging grating at $Z_T/4$ is used as the moving grating, then the signal period at the detector is also reduced by a factor of 2. If both the static index grating and the moving scale grating have a square pattern and the static grating is imaged sharply onto the moving grating, then the detected intensity will vary triangularly with displacement.

For accurate detection using phase quadrature interpolation, the optical system should be designed to generate a sinusoidal signal. The captured diffraction orders of the imaging system, the spectral width, the size and shape of the source and the distance between the gratings affect the sharpness of the imaged grating. With a proper encoder design a grating can be imaged as sinusoidal image such that, effectively, a sinusoidal modulation of the detector signal is obtained (Dürschmid, 1993). To obtain various phase signals in parallel from the same scale, the reticle mask is equipped with multiple grating areas that have predefined offsets. In the case of an incremental encoder,


S_0
S_{90}
S_{180}
S_{270}
4x photo detector
Index grating
Scale grating
$\Delta x = 0$
$\Delta x = \frac{1}{4}p$
$\Delta x = \frac{1}{2}p$
$\Delta x = \frac{3}{4}p$
50%
0%
S_0
S_{180}
Scale displacement Δx
50%
0%
S_{90}
S_{270}
p
Scale displacement Δx
Absolute reference mark

9.22 Encoder with four-phase zones with 90° phase offsets. For signal robustness, the phase quadrature signal is obtained by taking the difference signals s_0–s_{180} and s_{90}–s_{270}.

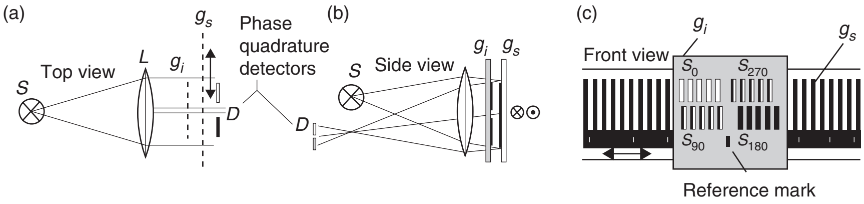

S_0
S_{90}
S_{180}
S_{270}
Array of photo detectors
Index grating
Scale grating

9.23 Optical encoder with an index grating and a scale grating with a small pitch difference.

a second head would read from the same moving grating but be displaced by $n + 1/4\cdot$ pitch ($n \in \mathbb{N}$) in order to produce a second signal with a phase offset of 90°. In practice, often four-phase quadrature detectors signals are generated. By taking the difference of signals that have an opposite phase ($s_0 - s_{180}$) and ($s_{90} - s_{270}$) the sensitivity for transmission losses of the electrical signal, causing DC offset variations, can be reduced (see Fig. 9.22).

As an alternative to using a single detector for each phase zone, the grating pitch of the index grating may be designed to have a pitch that is slightly different from the pitch of the scale grating. A segmented detector with binned pixels or a line camera is used to capture the moiré pattern that is formed by the light that has interacted with both the index grating and the scale grating (Fig. 9.23).

This encoder mask layout also generates directional information and has an improved robustness against local obscuration by dirt spots, as the effect of dirt is more equally distributed over all phase signals.



9.24 (a) Transmission type encoder, (b) reflective type encoder, (c) front view, indicating phase zones.

Note that imaging encoders may be designed in transmission (Fig. 9.24a) or reflection (Fig. 9.24b). Transmission scales are typically etched chrome on glass gratings, while reflective scales often are gold on metal. Reflective scales are often preferred as they are relatively easy to integrate into a mechanical design. Especially for long stroke actuation, the reflective scales can be easily mounted to any rigid surface and do not need to be self-supporting.

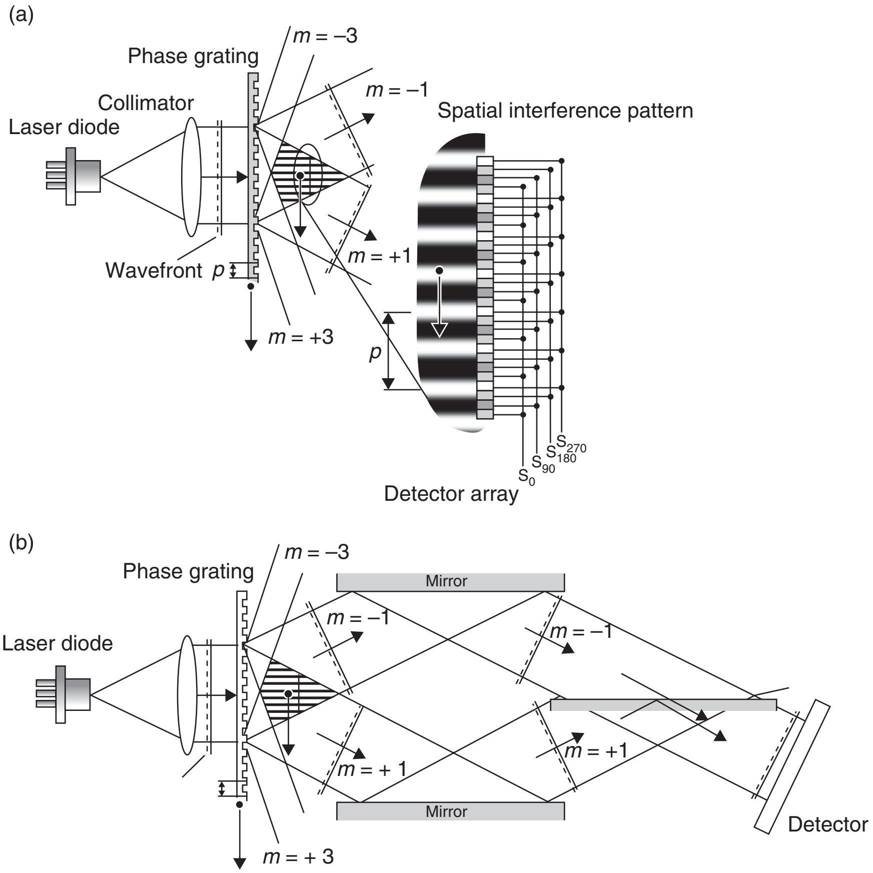
9.4.2 Interferential encoders

The resolution of an optical encoder scales directly with the scale period. With a scale period smaller than 10 μm – for example as small as 0.5 μm – diffraction effects become dominant and can be exploited in interferential encoders, also known as 'grating interferometers', to achieve the highest levels of precision. The detector signals are sinusoidal with grating displacement and largely free of harmonics. With proper electronics these signals can be interpolated more than 1000 times, practically being limited by the signal noise.

The optical path lengths of the grating interferometer arms are matched and short compared with those of displacement interferometers. This is in order to reduce the sensitivity to wavelength variations as caused by refractive index variations or source frequency variations.

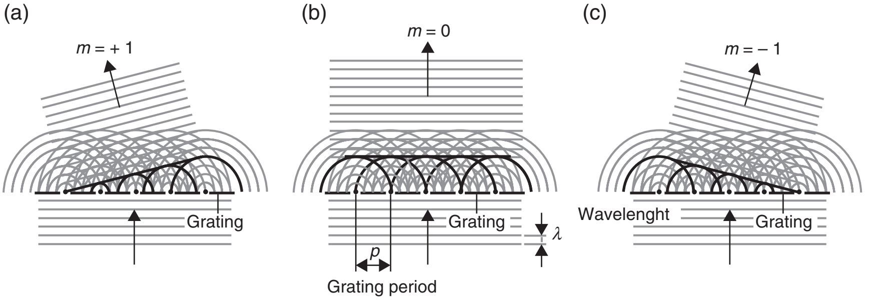
Interferential grating encoders use coherent light-emitting diodes or semiconductor lasers to generate a coherent and collimated light beam. In the example in Fig. 9.25, the light beam is split into diffraction orders (m) by the scale grating with a pitch p. Typically the –1 and +1 diffraction orders are selected to form the interferometer arms. When the +1 and the –1 diffracted orders are combined to interfere at a photodetector, a sinusoidal signal with a signal period of half the grating pitch is created:

$$\Delta\phi(\Delta x) = \Delta\phi_{m=+1} - \Delta\phi_{m=-1} = 2\pi(+1)\frac{\Delta x}{p} - 2\pi(-1)\frac{\Delta x}{p} = 4\pi\frac{\Delta x}{p} \qquad [9.24]$$


(a)
Phase grating
m = –3
m = –1
Collimator
Laser diode
Spatial interference pattern
m = +1
Wavefront
p
m = +3
p
S270
S180
S90
S0
Detector array
(b)
Phase grating
m = –3
Mirror
Laser diode
m = –1
m = –1
m = + 1
m = +1
Mirror
Detector
m = + 3

9.25 Basic principle of interferential grating encoder. (a) Direct detection of the spatial phase of a spatial interference pattern of the +1 and –1 diffracted order, (b) detection of the interference signal after both interferometer arms are made parallel.

As a result, for in-plane displacements the signal period is directly related to the grating pitch of the scale and not related to the wavelength of the light source. In Fig. 9.25, two schematic representations of interferential encoders in their simplest form are presented. In Fig. 9.25a, the beams of the +1 and –1 diffracted orders are made to interfere when recombined at a photodetector to form a spatial sinusoidal pattern. The period of the spatial interference pattern is equal to half the grating pitch and the lines are parallel to the angle bisector of the angle between the two diffracted orders. In Fig. 9.25b, the beams of the +1 and –1 diffracted orders are made parallel such that no spatial pattern emerges and the signal that varies



9.26 (a–c) Light diffraction of a coherent light beam by an amplitude grating resulting in multiple diffracted beams.

periodically with grating displacement can be detected by a single photosensitive detector.

Diffraction physics

In Fig. 9.26, the interaction between a periodic grating structure and a coherent light beam is shown. The Huygens-Fresnel principle is used for a simplified understanding when a beam that is diffracted by a grating may be thought of as an array of coherent point sources, each emitting a spherical wave. In specific propagation directions, the spherical waves from all point sources interfere constructively and will propagate as a plane wavefront. This diffracted wavefront is indicated graphically by the straight lines that connect the spherical wavefronts of neighboring slits.

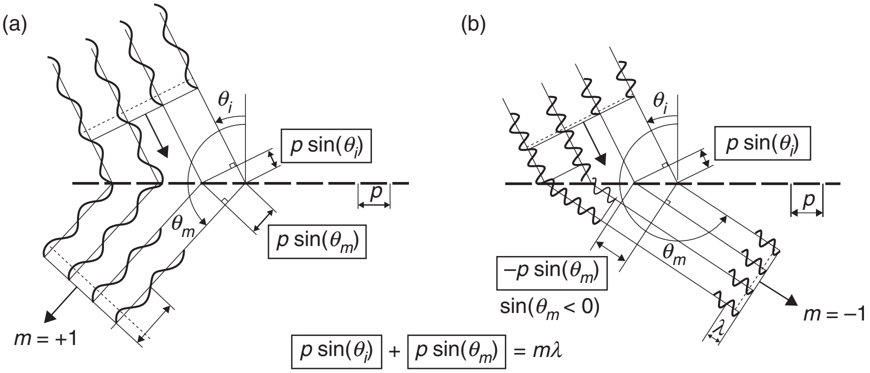
For constructive interference to occur, the optical path length from an input wavefront to an output wavefront must differ by integer multiples of λ for adjacent slits of the scale grating (see Fig. 9.27). This integer multiple m is called the order of diffraction.

The relationship as visualized in Fig. 9.27 between the entrance angle θ_i and the output angle θ_m of a diffracted beam of order m, wavelength λ and grating pitch p is then described by the general grating equation:

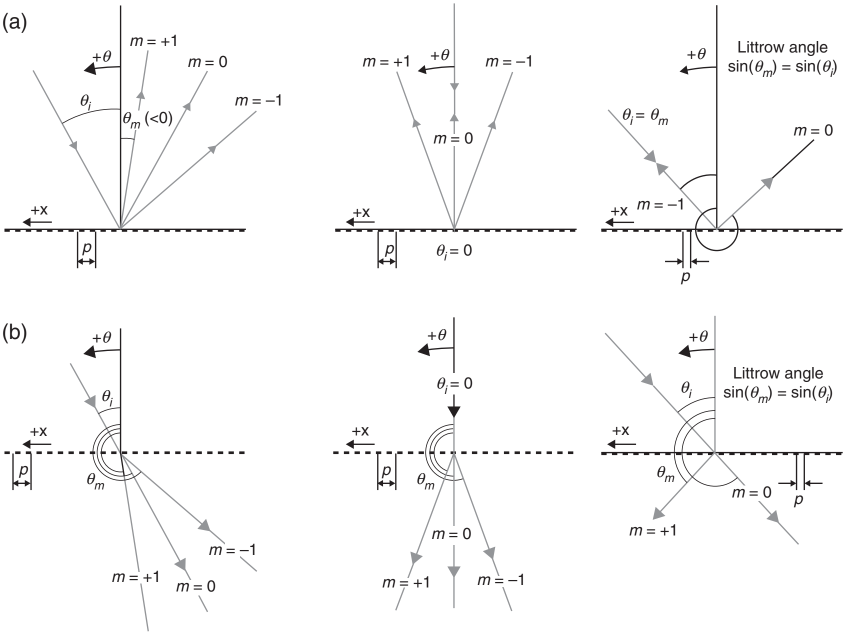
$$\sin(\theta_i) + \sin(\theta_m) = \frac{m\lambda}{p} \qquad [9.25]$$

When the input beam is perpendicular to the grating ($\theta_i = 0$), this equation simplifies to:

$$\sin(\theta_m) = \frac{m\lambda}{p} \qquad [9.26]$$


(a)
(b)
θ_i
$p\sin(\theta_i)$
p
θ_m
$p\sin(\theta_m)$
$m = +1$
θ_i
$p\sin(\theta_i)$
p
θ_m
$-p\sin(\theta_m)$
$\sin(\theta_m < 0)$
λ
$m = -1$
$\boxed{p\sin(\theta_i)} + \boxed{p\sin(\theta_m)} = m\lambda$

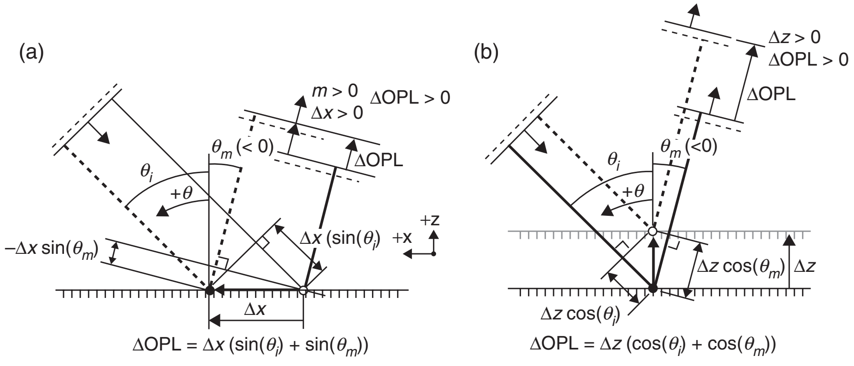
9.27 Derivation of the grating equation. Going from left to right 1λ. (a) $m = +1$, (b) $m = -1$.


(a)
$+\theta$
$m = +1$
$m = 0$
$m = -1$
θ_i
θ_m (<0)
+x
p
$m = +1$
$+\theta$
$m = -1$
$m = 0$
+x
p
$\theta_i = 0$
$+\theta$
Littrow angle
$\sin(\theta_m) = \sin(\theta_i)$
$\theta_i = \theta_m$
$m = 0$
+x
$m = -1$
p
(b)
$+\theta$
θ_i
+x
p
θ_m
$m = -1$
$m = +1$
$m = 0$
$+\theta$
$\theta_i = 0$
+x
p
θ_m
$m = 0$
$m = +1$
$m = -1$
$+\theta$
Littrow angle
$\sin(\theta_m) = \sin(\theta_i)$
θ_i
+x
θ_m
$m = 0$
p
$m = +1$

9.28 Reflective (a) and transmissive (b) diffraction angles according the general grating equation. Left: General. Center: Normal incidence ($\theta_i = 0$). Right: Littrow angle ($\sin(\theta)_m) = (\sin(\theta)_i$).

The grating equation applies for diffraction by both transmissive and reflective gratings as can be seen in the examples in Fig. 9.28.

Diffraction at the Littrow angle is considered a special case where, for a reflected beam, the diffracted beam travels in the direction opposite to the input beam.


(a)
$m > 0$
$\Delta x > 0$
ΔOPL > 0
θ_m (< 0)
ΔOPL
θ_i
$+\theta$
+z
Δx (sin(θ_i)
+x
$-\Delta x$ sin(θ_m)
Δx
ΔOPL = Δx (sin(θ_i) + sin(θ_m))
(b)
$\Delta z > 0$
ΔOPL > 0
ΔOPL
θ_i
θ_m (<0)
$+\theta$
Δz cos(θ_m)
Δz
Δz cos(θ_i)
ΔOPL = Δz (cos(θ_i) + cos(θ_m))

9.29 Phase sensitivity of a diffracted wavefront. For (a) in-plane grating displacement and (b) out-of-plane grating displacement.

Sensitivities

To understand the phase shift sensitivity as a function of the diffraction angles, let us consider the situation of Fig. 9.29. Here, the optical path length differences are plotted for a diffracted beam under the influence of an in-plane and an out-of-plane displacement.

By observing the change in optical path length from an input wavefront to an output wavefront for different grating positions, the phase sensitivity for both the in-plane displacement (Δx in Fig. 9.29) and out-of-plane (Δz in Fig. 9.29) displacement of the grating can be derived geometrically. The beams depicted by the solid line and the dashed line in Fig. 9.29 are of equal length. Using the length difference of the entrance beam ($\Delta x \sin(\theta_i)$) and the length difference of the reflected beam ($\Delta x \sin(\theta_m)$), the optical path length difference can be calculated:

$$\frac{\Delta\text{OPL}}{\Delta x} = \left(\sin(\theta_i) + \sin(\theta_m)\right) \qquad [9.27]$$

$$\frac{\Delta\text{OPL}}{\Delta z} = \left(\cos(\theta_i) + \cos(\theta_m)\right) \qquad [9.28]$$

As, finally, a phase difference between two interferometer arms is measured, the sensitivities for the diffraction on a moving scale are better expressed in phase terms:

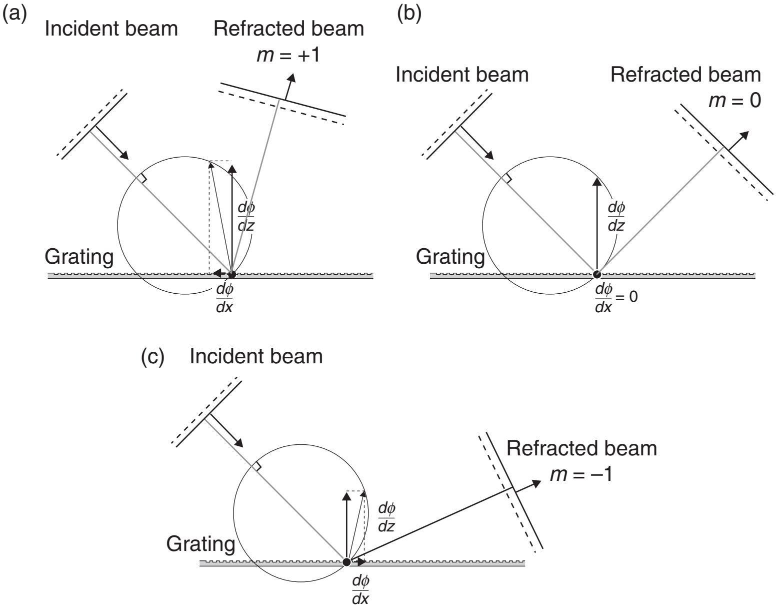
$$\frac{d\phi}{dx} = \frac{2\pi}{\lambda} \cdot \frac{\Delta\text{OPL}}{\Delta x} = \frac{2\pi}{\lambda}\left(\sin(\theta_i) + \sin(\theta_m)\right) = 2\pi\frac{m}{p} \qquad [9.29]$$

$$\frac{d\phi}{dz}=\frac{2\pi}{\lambda}.\frac{\Delta \mathrm{OPL}}{\Delta z}=\frac{2\pi}{\lambda}\left(\cos(\theta_i)+\cos(\theta_m)\right) \tag{9.30}$$

The resulting sensitivity vector for a single diffracted beam can be defined in complex notation as:

$$\vec{S}=\frac{d\phi}{dx}+1i\frac{d\phi}{dz} \tag{9.31}$$

The sensitivity vector of an interferential encoder sensor is obtained by taking the difference between the sensitivity vectors of each interferometer arm. Examples of a sensitivity vector for reflective and transmissive scale interactions at various output angles are shown in Figs 9.30 and 9.31. In the case of a multi-pass encoder, the sensitivity for each interferometer arm is obtained by summing the sensitivities of every interaction between the beam in the interferometer arm and the moving scale grating. It can be seen in Equation [9.29] that the in-plane sensitivity $d\phi/dx$ only depends on the scale pitch and that there is no wavelength dependency. The minimum z-sensitivity of an interferometer arm for out-of-plane displacements


(a)
Incident beam
Refracted beam
m = +1
dϕ/dz
Grating
dϕ/dx
(b)
Incident beam
Refracted beam
m = 0
dϕ/dz
Grating
dϕ/dx = 0
(c)
Incident beam
Refracted beam
m = −1
dϕ/dz
Grating
dϕ/dx

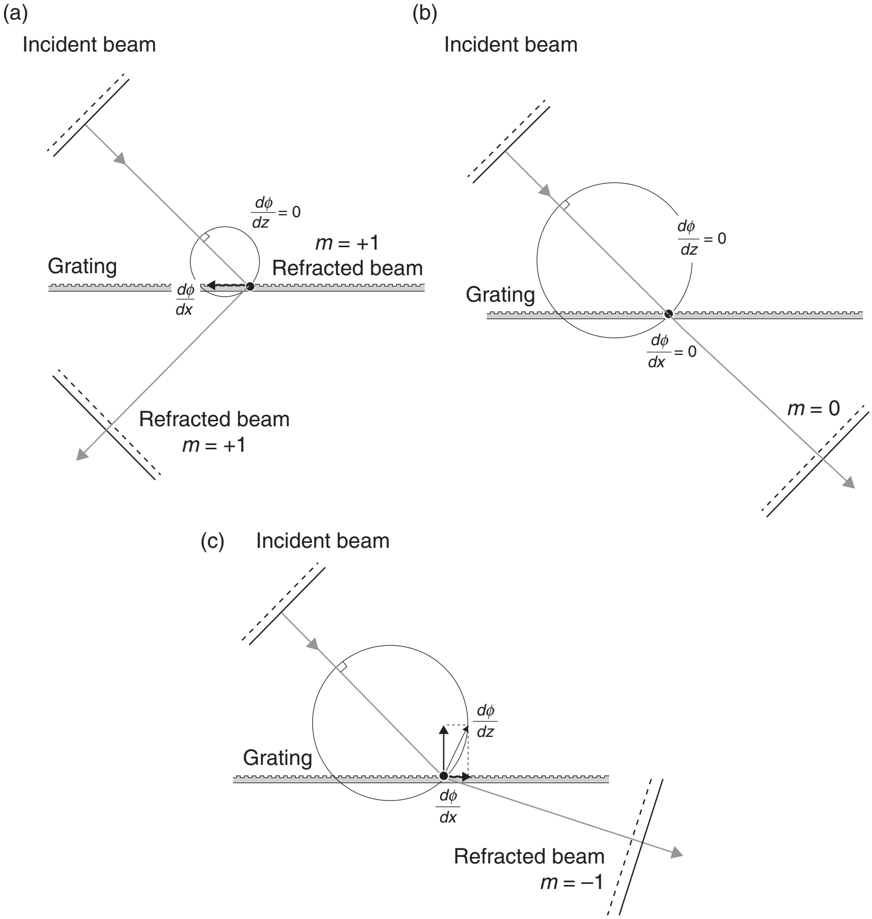
9.30 Example of the sensitivity vectors of a reflective scale grating. for the diffraction orders (a) $m = -1$, (b) $m = 0$, (c) $m = +1$.

(Equation [9.30]) is found for transmissive diffraction at the largest possible Littrow angle. A large Littrow angle is obtained for a long wavelength and/or a small scale pitch (see Figs 9.31a and 9.32b).

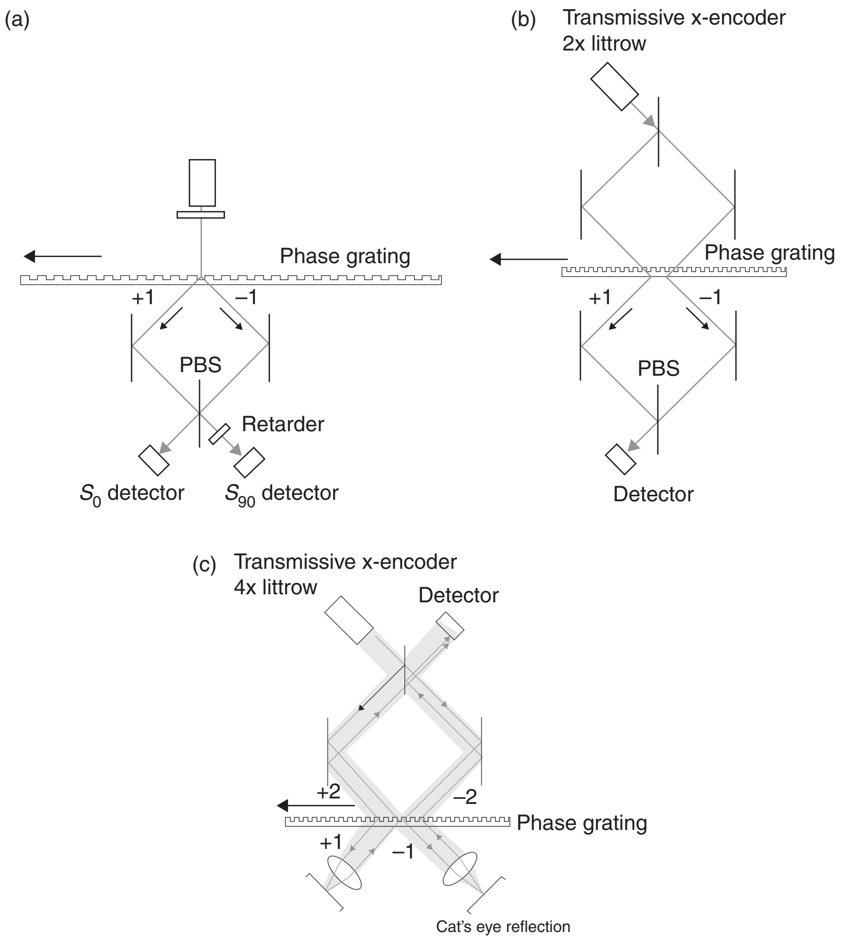
A symmetric design of the optical paths of the interferometer arms will balance the sensitivity for out-of-plane displacements, while the sensitivity for in-plane grating displacement is enhanced by interfering beams of opposite diffraction orders.

Schematic setups

Many different designs of high-resolution optical encoders can be conceived. Figures 9.32 and 9.33 show some typical configurations. The ratio


(a)
Incident beam
$\frac{d\phi}{dz} = 0$
$m = +1$
Refracted beam
Grating
$\frac{d\phi}{dx}$
Refracted beam
$m = +1$
(b)
Incident beam
$\frac{d\phi}{dz} = 0$
Grating
$\frac{d\phi}{dx} = 0$
$m = 0$
(c)
Incident beam
$\frac{d\phi}{dz}$
Grating
$\frac{d\phi}{dx}$
Refracted beam
$m = -1$

9.31 Example of the (normalized) sensitivity vectors of a transmissive scale grating. For the diffraction orders (a) $m = -1$, (b) $m = 0$, (c) $m = +1$.


(a)
Phase grating
+1
−1
PBS
Retarder
S_0 detector
S_{90} detector
(b) Transmissive x-encoder
2x littrow
Phase grating
+1
−1
PBS
Detector
(c) Transmissive x-encoder
4x littrow
Detector
+2
−2
Phase grating
+1
−1
Cat's eye reflection

9.32 Schematic presentation of transmissive diffraction encoders and the sensitivities. (a) Basic, (b) littrow configuration for small pitch grating, (c) double-pass for increased resolution.

between the pitch of the grating scale and the signal period depends on the selected diffraction orders of the interfering beams and the number of times a beam has been diffracted by the scale grating. If the light beams of the interferometer arms are formed by the +1 and −1 diffraction orders and only diffracted once by the scale grating, then the resulting signal period will be half the grating pitch. In case of a double-pass grating interferometer, the signal pitch will be reduced to one-quarter of the grating pitch. In most designs a phase grating is used instead of an amplitude grating, such that the 0th order is suppressed and the optical power is directed toward the +1 and −1 diffraction orders.

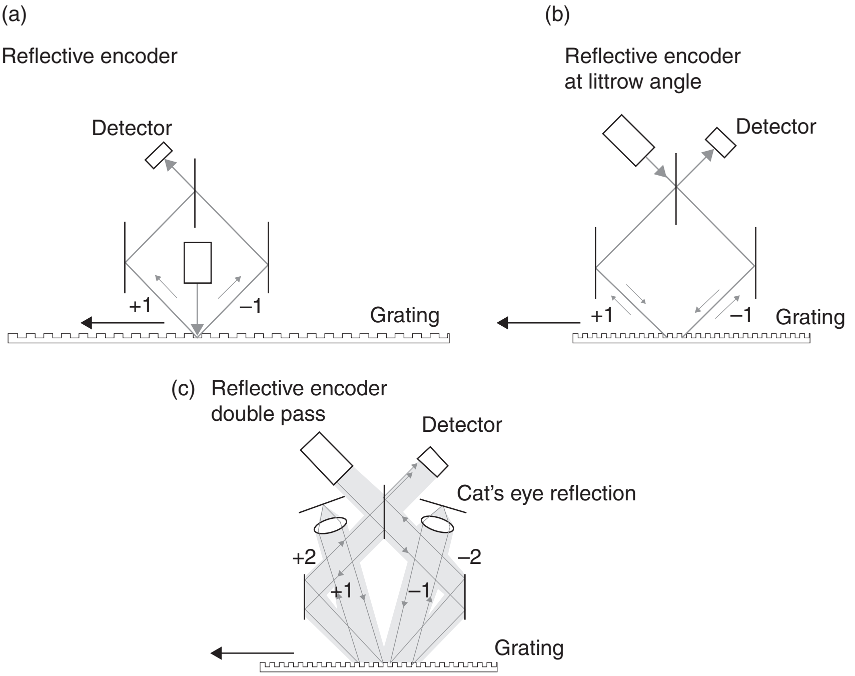
In Fig. 9.32a, a most basic interferential grating encoder is shown. A coherent beam is diffracted into a +1 and −1 diffraction order by a transmissive phase grating, forming the two interferometer arms with opposite phase offsets as a function of grating displacement. When the two interferometer arms are combined, a sinusoidal interference signal as a function of grating displacement is obtained with a period of half the grating pitch. For the design of a high-resolution optical encoder, it is beneficial to have a high in-plane sensitivity which is obtained by choosing a small grating pitch. At the same time, the out-of-plane senstivity is preferably reduced by choosing large diffraction angles and the longest possible wavelength. These benefits are combined in the example in Fig. 9.32b, where the beams are diffracted under the Littrow angle ($\sin(\theta_i) = \sin(\theta_m)$). The example in Fig. 9.32c shows a schematic presentation of a double-pass transmission grating encoder. Here, the phase shift that is introduced when the beams are first diffracted by the moving scale grating, is doubled during the diffraction that occurs in the second pass. With one interferometer arm receiving a positive phase shift twice and the other interferometer arm receiving a negative phase shift twice, a total of four-phase cycles is obtained for every grating displacement equal to the grating pitch.

The interferential encoder concepts in Fig. 9.32, using transmissive scale gratings, are depicted as reflective interferential grating encoders in Fig. 9.33. The cat's eye retroreflection in the schematic representation of a double-pass reflective encoder layout of Fig. 9.33c improves tolerances for grating tilts and makes the optical path length from the source to the detector constant across the beam, lowering the requirements on the source coherence.

Phase detection

In most cases the encoder will be designed such that the interferometer arms have an orthogonal polarization state and the detector can be formed by a homodyne phase analyzer to determine the phase difference between two interferometer arms (Fig. 9.34).

In a typical homodyne phase analyzer, the polarization states of the interferometer arms are made circular and counter-rotating using a quarter-wave retarder plate oriented at an angle of 45°. The sum of the two circularly polarized beams is then a linearly polarized beam with a polarization angle that is equal to half the phase difference $\Delta\phi / 2$. The phase quadrature signals (from which $\Delta\phi$ is derived using an arctangent calculation) are obtained by splitting the linearly polarized beam, and analyze the intensity after passing the light through analyzing polarizers with different angular alignment. By splitting the beam and passing each beam through a polarizer at 0° and 45°, two phase quadrature signals are generated with a relative phase offset of 90°. When plotted against each other, the measured phase $\Delta\phi$ can be read

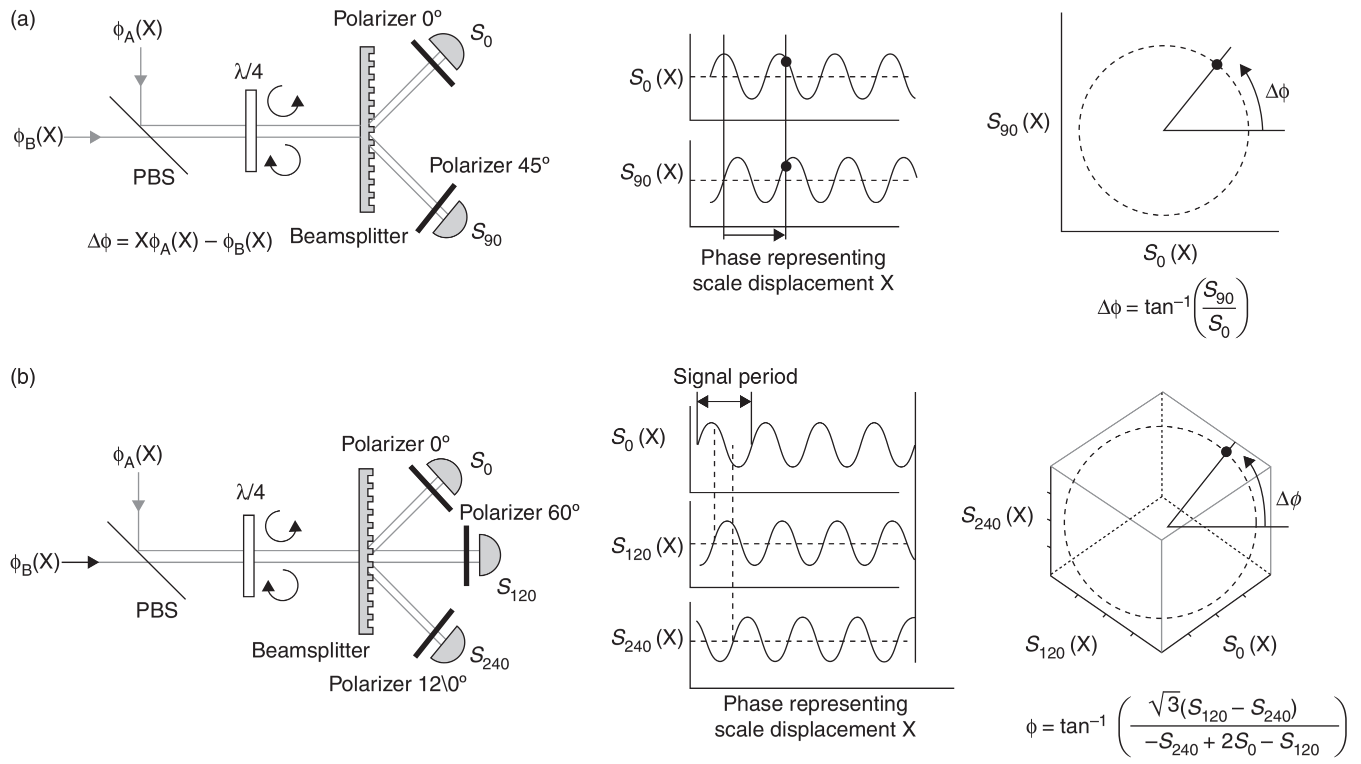

(a)
Reflective encoder
Detector
+1
−1
Grating
(b)
Reflective encoder at littrow angle
Detector
+1
−1
Grating
(c)
Reflective encoder double pass
Detector
Cat's eye reflection
+2
−2
+1
−1
Grating

9.33 Schematic presentation of interferential encoders using reflective scales. (a) Basic, (b) littrow configuration for small pitch grating, (c) double-pass for increased resolution.

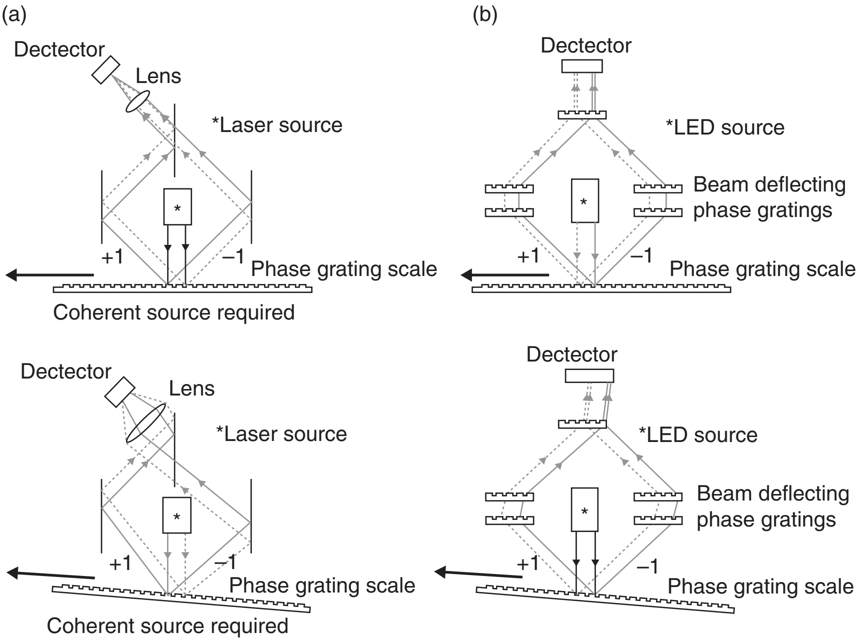
out in a 2D Lissajous plot. The beam may be split in three to generate phase signals with an offset of 0°, 120° and 240°, allowing for a more robust phase calculation and compensation of periodic errors. For three phase signals with 120° phase offsets, the measured phase can be read out in a 2D projection of a 3D Lissajous plot (see Fig. 9.34).

Tilt sensitivity

For a robust design, the tilt sensitivity must be designed such that both interferometer arms are parallel at the detector. To reduce the sensitivity to wavefront deviations, the design should also prevent a large beam walk off between the interfering arms. To allow the use of a short coherence source, the optical paths from the source to the detector should be identical across the beam diameter. As can be seen in Fig. 9.35, typically good design characteristics can be obtained when using grating beam splitters and grating based beam deflectors.


(a)
$\phi_A(X)$
$\phi_B(X)$
PBS
$\lambda/4$
Polarizer 0°
S_0
Polarizer 45°
S_{90}
Beamsplitter
$\Delta\phi = X\phi_A(X) - \phi_B(X)$
S_0 (X)
S_{90} (X)
Phase representing scale displacement X
S_{90} (X)
S_0 (X)
$\Delta\phi$
$\Delta\phi = \tan^{-1}\left(\frac{S_{90}}{S_0}\right)$
(b)
$\phi_A(X)$
$\phi_B(X)$
PBS
$\lambda/4$
Polarizer 0°
S_0
Polarizer 60°
S_{120}
S_{240}
Beamsplitter
Polarizer 12\0°
Signal period
S_0 (X)
S_{120} (X)
S_{240} (X)
Phase representing scale displacement X
S_{240} (X)
S_{120} (X)
S_0 (X)
$\Delta\phi$
$\phi = \tan^{-1}\left(\frac{\sqrt{3}(S_{120} - S_{240})}{-S_{240} + 2S_0 - S_{120}}\right)$

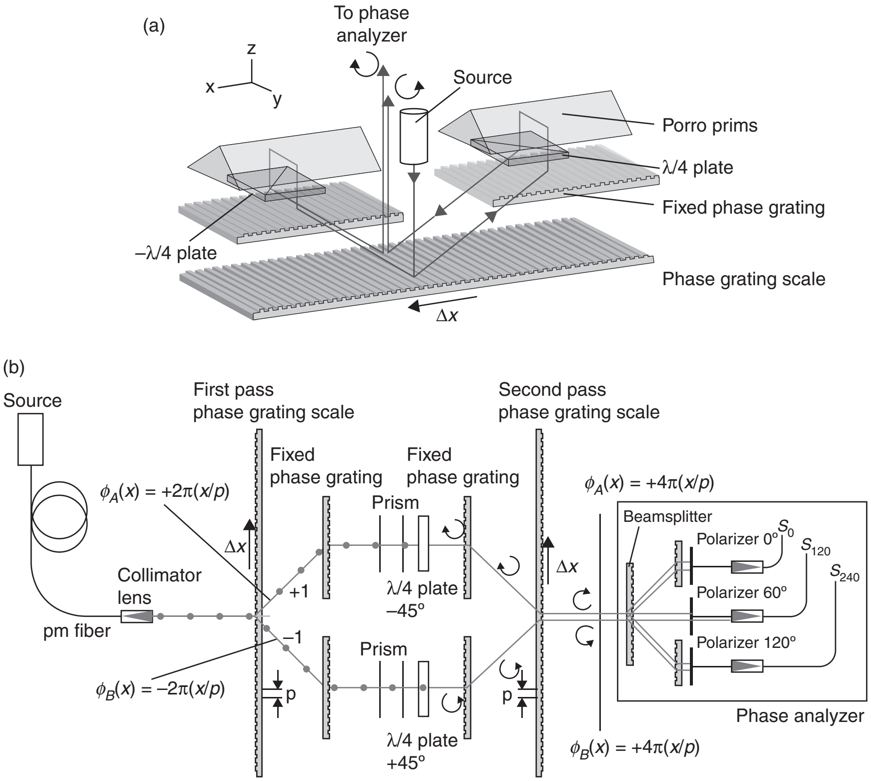
9.34 Homodyne phase detector. For creating phase quadrature signals representing the phase difference ($\Delta\phi = \phi_A - \phi_B$) between the orthogonally polarized interferometer arms of the encoder: (a) for two phase signals with 90° offsets, the measured phase can be read out in a 2D Lissajous plot, (b) for three phase signals with 120° phase offsets the measured phase can be read out in a 2D projection of a 3D Lissajous plot.


(a)
Dectector
Lens
*Laser source
+1
−1
Phase grating scale
Coherent source required
(b)
Dectector
*LED source
Beam deflecting phase gratings
+1
−1
Phase grating scale
Dectector
Lens
*Laser source
+1
−1
Phase grating scale
Coherent source required
Dectector
*LED source
Beam deflecting phase gratings
+1
−1
Phase grating scale

9.35 (a) Design requiring spatially coherent source and with low tilt tolerance, (b) improved design.

Practical example

Figure 9.36a shows a double-pass interferential grating encoder. Here, a collimated linearly polarized beam is delivered via a polarization maintaining single mode fiber, and directed towards the grating scale. The grating scale is a reflective phase grating with a groove depth of $\lambda/4$, designed to suppress the zeroth diffraction order while boosting the intensities of the first-order diffracted beams. The beam is diffracted by the grating scale into a positive and a negative diffraction angle (m =+1 and m = −1), giving each refracted beam an opposite phase shift ($m \cdot 2\pi\Delta x/p$). Each of the interferometer arms is diffracted towards a Porro prism by a static grating, which gives the beam a shift in y-direction. The linear polarization states of the beams in the interferometer arms are then made circular and counter-rotating using $\lambda/4$ retarder plates with an angular alignment of −45° and +45°. On return, the static phase grating directs the beam towards the scale grating where the beams of both interferometer arms are recombined to receive an additional phase shift upon diffraction by the grating scale. The phase difference between the interferometer arms ($\Delta\phi$) is now equal to $8\pi\Delta x/p$, meaning that four-phase cycles are detected for a grating displacement Δx that is equal to the scale pitch p. With the polarization state of the interferometer arms being circular and counter-rotating,


(a)
To phase analyzer
z
x
y
Source
Porro prims
λ/4 plate
Fixed phase grating
−λ/4 plate
Phase grating scale
Δx
(b)
Source
First pass phase grating scale
Second pass phase grating scale
Fixed phase grating
Fixed phase grating
$\phi_A(x) = +2\pi(x/p)$
Prism
$\phi_A(x) = +4\pi(x/p)$
Beamsplitter
Δx
Δx
Polarizer 0°
S_0
S_{120}
S_{240}
Collimator lens
+1
λ/4 plate −45°
Polarizer 60°
pm fiber
−1
Prism
Polarizer 120°
$\phi_B(x) = -2\pi(x/p)$
p
p
Phase analyzer
λ/4 plate +45°
$\phi_B(x) = +4\pi(x/p)$

9.36 (a, b) Double-pass interferential grating encoder with a signal period of one-quarter of the grating period (Holzapfel and Linnemann, 2006).

the phase difference between the interferometer arms is detected using a homodyne phase analyzer as depicted in Fig. 9.34.

9.5 Design considerations

Highly stable optical encoders require a thermally and dynamically stable design as well as high-quality optics and electronics. The design of an encoder should reduce the sensitivity to manufacturing and mounting tolerances, as well as reduce the influence of environmental conditions like temperature gradients and temperature offsets. Errors should be counterbalanced by design where possible.

9.5.1 Stability

The optical path lengths of the grating interferometer arms are matched and short compared with those of displacement interferometers; this is in order to reduce the sensitivity to wavelength variations as caused by refractive index variations or source frequency variations. For applications requiring sub-nm stability, a thermally stable design is obtained by using compact and symmetric design with a short distance between the scale and the reading head, and by using low-thermal expansion materials. The light source and detector signals may be fiber-coupled in order to reduce the heat dissipation in the encoder head. This is typically important in vacuum applications where heat cannot be dissipated by means of natural convection. If heat is dissipated inside the sensor, there should be a good thermal interface with its surrounding. For a deterministic behavior, the reading head should be mounted with kinematic mounts and the scale should be allowed to expand freely to make sure that no forces are exerted on the scale as a result of thermal expansion of the carrier material (Breyer *et al.*, 1991).

9.5.2 Grating scale errors

Measurement uncertainty is further determined by the manufacturing uncertainty of the grating scale. Low- and mid-frequency deviations must be calibrated against an interferometer in vacuum or in air (Maeda, 2001; Holzapfel, 2008). When measured in air the refractive index variations due to pressure, temperature and humidity variations are monitored and compensated for using the Edlén equation (Edlén, 1966; Birch, 1993; Bönsch and Potulski, 1998).

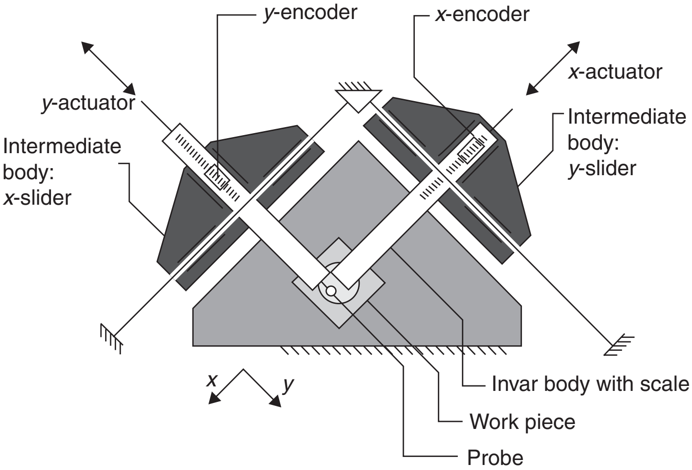
9.5.3 Periodic errors

Where imaging encoders must be carefully designed to obtain a purely sinusoidal pattern, the scanning signals of interferential encoders are always sinusoidal and largely free of harmonics. Periodic errors (sometimes referred to as ‘cyclic errors’) may still be present as a result of variations in light intensity, polarization mixing of non-ideal optics, mechanical misalignment of components in the reading head and electronic gain errors. Most of the periodic deviations may be compensated for by a phase error lookup table or Heydemann correction (Heydemann, 1981). However, there are some factors that are difficult to correct for fully, as the periodic deviations may be speed, position (scale alignment errors/dirt on the scale) and time (laser power variations, mechanical stability) dependent. With proper electronics the encoder signals can be interpolated as much as a few thousand times, practically being limited by signal noise. The impact of periodic errors may be reduced by adopting inline software correction and phase gratings to suppress the effect of unused diffraction orders.

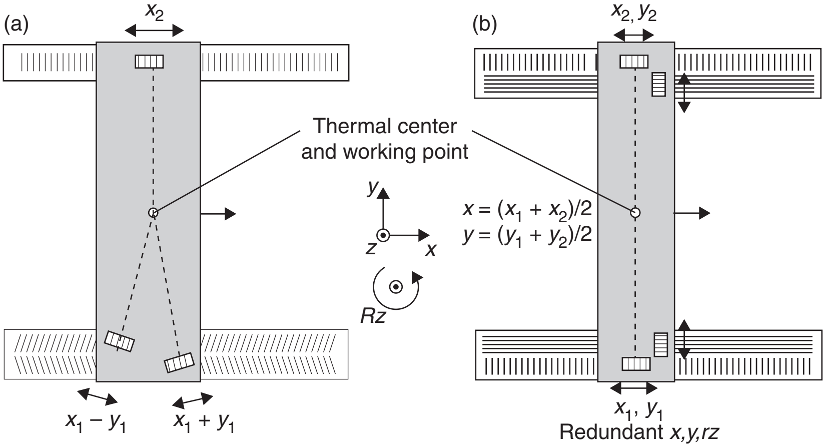
9.5.4 Abbe correction

According to Abbe's principle, the measured point on the object and the effective measuring point of the encoder should be aligned along the measuring direction. An example of an encoder-based nano-CMM working according to the Abbe principle in the *x*,*y* plane is presented in Fig. 9.37 (Vermeulen, 1999; VanSeggelen, 2007). Here, the measurement directions of the *x* and *y* scales are aligned to intersect at the work point. Rather than moving orthogonally and independently of each other, as is the case for most CMMs, the *x* and *y* axes are connected together at right angles and move as a single unit. To follow the motion of the unit in the direction orthogonal to the measuring directions of the encoder reading heads, the reading head of the *x*-encoder is mounted on an intermediate body (*y*-slider) and the *y*-encoder is mounted on an *x*-slider. Here, both encoder scales are aligned to intersect at the work point, where in this case a tactile stylus probe is mounted, thus eliminating Abbe errors in the 2D plane.

In configurations that do not strickly adhere to the Abbe principle of alignment, the Abbe error may be reduced to a practical minimum by, for example, a phantom scale arrangement or a cross-grid scale. With a phantom scale arrangement, additional stage rotation is measured using another scale with a different Abbe offset, such that the Abbe error as induced by stage rotation can be determined and compensated for mathematically (see also Fig. 9.38). Alternatively, a 2DOF (degrees-of-freedom) grid encoder uses a cross-grid and may be used to measure close to the work point to minimize the Abbe errors.



9.37 Encoder-based system with scales oriented to be collinear with the work point to eliminate Abbe errors (Vermeulen, 1999; Van Seggelen, 2007).



9.38 (a) *X,Y, Rz* encoder configuration with a thermal center. 3DOF scale arrangement with grating lines pointing towards the work point in the thermal center to reduce sensitivity for thermal expansion; additional rz measurement allows for a mathematical Abbe offset compensation. (b) Redundant *x,y,rz* systems allowing for a mathematical Abbe offset and thermal expansion compensation (Kuhn, 2010).

9.5.5 Thermal expansion

If the effect of thermal expansion cannot be ignored and it is not possible or impractical to position an encoder directly at the work point and thermal center, then a single 6DOF sensor positioned outside the thermal center will suffer from Abbe measurement errors, and no correction for thermal expansion between the work point and the measurement positions can be applied. It may then be necessary to add measurement redundancy to be able to measure and correct for thermal expansion effects.

Also, multiple encoders may be used in a scale layout such that the sensitivity to thermal expansion effects is minimized, as is depicted in the example of Fig. 9.38a. Here, a scale configuration is shown for measuring the *x*-displacement and guide errors in *y* and *rz* direction. The *rz* measurement allows for Abbe arm correction at the work point. For each encoder, the grating lines of the scale are pointing towards the thermal center of the stage, such that the measured position in the thermal center position is not affected by uniform thermal expansion of the stage. The combined measurement sensitivity for each degree of freedom may be tuned by choosing appropriate grid orientations and encoder read head positions. In Fig. 9.38b, *x, y* and *rz* can also be measured and corrected for Abbe offsets. Here, the thermal expansion of the scale can be deduced from the difference in the redundant *y*-measurements and corrected for mathematically. Note that with the use of 2DOF 'XZ'-encoders like presented in Fig. 9.40, the configuration of Fig. 9.38 may be used to measure 6 degrees of freedom of a moving stage.

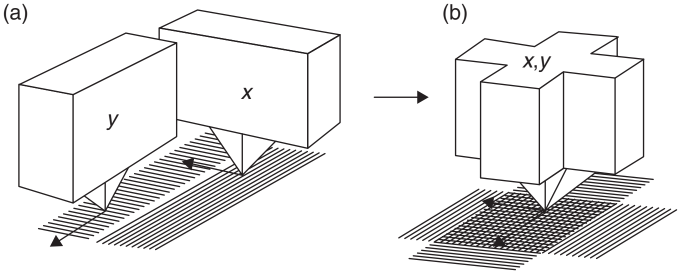
9.5.6 Multi axis encoder systems

Complete 6DOF stage position measurements using optical encoders is obtained by combining multiple measurement axes in a single encoder package (Sandoz, 2005; Fan *et al.*, 2008) and/or by combining multiple separate optical encoder systems which are strategically positioned to obtain desired measurement sensitivity for each measured degree of freedom. Sensor and scale arrangements may be chosen such that potential error sources as thermal expansion or guiding errors do not affect the measurement, or can be separated from the measurement signal mathematically by using redundant measurement axes.

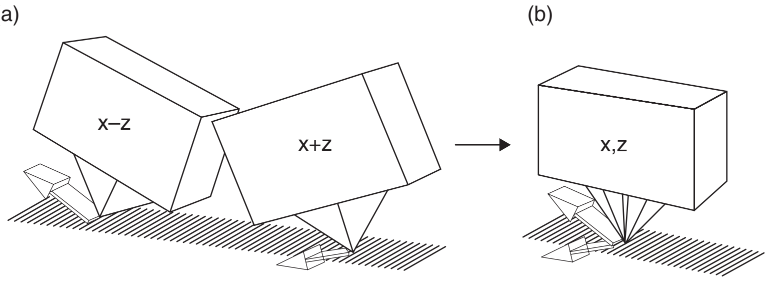
An XY encoder may have a scale with separate zones for x and y gratings. In Fig. 9.39a, a long stroke main grating is intended for displacement measurement, while the short stroke can be used for direct compensation of linear guide straightness. Yaw measurement becomes possible with an additional x-encoders positioned at a different y-position. For 2D contouring, a 2D cross-grating and an integrated XY encoder may be used (Fig. 9.39b). The cross-grid grating plate generates diffraction orders in both x origin and y directions (Fan *et al.*, 2008).

When an encoder is designed to be placed under an angle it becomes sensitive for both in-plane and out-of-plane motion (see Figs 9.40 and 9.41). The encoder signals $E_1(x, z)$ and $E_2(x, z)$ of two encoders with an opposite tilt are used to obtain an in-plane and out-of-plane displacement signal.

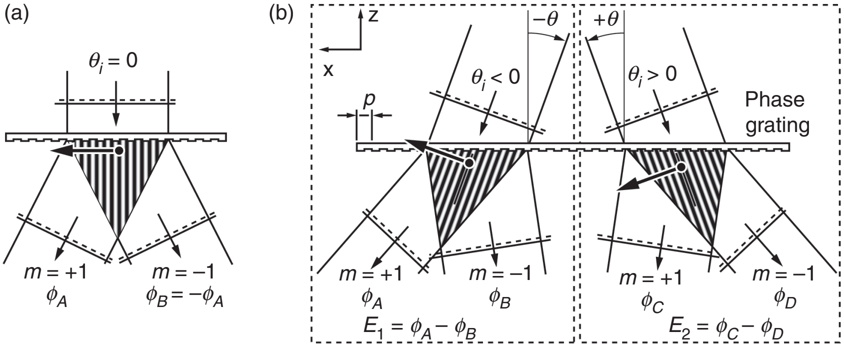
Given the wavelength λ, the grating pitch p and the input angles of the diffracted beams, the encoder signals $E_1(x, z)$ and $E_2(x, z)$ can be derived from the phase shift sensitivities of each of the four interferometer arms



9.39 (a) Separate X and Y-encoder (Kuhn, 2010), (b) XY cross-grid encoder with x and y-encoder enclosed in a single housing (Huber *et al.*, 1993).



9.40 (a, b) Simplified representation of an XZ encoder concept based on integration of two tilted encoder heads (Holzapfel, 2009).


(a)
$\theta_i = 0$
$m = +1$
ϕ_A
$m = -1$
$\phi_B = -\phi_A$
(b)
z
x
$-\theta$
$+\theta$
$\theta_i < 0$
$\theta_i > 0$
p
Phase grating
$m = +1$
ϕ_A
$m = -1$
ϕ_B
$E_1 = \phi_A - \phi_B$
$m = +1$
ϕ_C
$m = -1$
ϕ_D
$E_2 = \phi_C - \phi_D$

9.41 (a) Measurement direction for a single symmetric encoder $\cos(\theta_{m=+1}) = \cos(\theta_{m=-1})$, (b) effective measurement direction for two symmetric interferential encoders with non-normal incidence (transmissive type for easier representation). The vector sum or subtraction of both encoder signals make an X or Z measurement signal.

with relative phases ($\phi_{A,}\ \phi_{B,}\ \phi_{C,}\ \phi_D$) by using the grating Equations [9.1] and the derived senstivity Equations [9.5] and [9.6]:

$$E_1(x,z) = \phi_A(x,z) - \phi_B(x,z) = \frac{4\pi}{p}\cdot\Delta x - \frac{2\pi}{\lambda}\left(\cos(\theta_{m\theta_A}) - \cos(\theta_{m\theta_B})\right)\cdot\Delta z \quad [9.32]$$

$$E_2(x,z) = \phi_C(x,z) - \phi_D(x,z) = \frac{4\pi}{p}\cdot\Delta x + \frac{2\pi}{\lambda}\left(\cos(\theta_{m\theta_D}) - \cos(\theta_{m\theta_C})\right)\cdot\Delta z$$

From the symmetry in Fig. 9.41b, it follows that $\cos(\theta_{m\theta A}) = \cos(\theta_{m\theta D})$ and $\cos(\theta_{m\theta C}) = \cos(\theta_{m\theta B})$. The in-plane and out-of-plane displacements are then derived by combining the (single pass) encoders signals $E_1(x, z)$ and $E_2(x, z)$:

$$
\begin{aligned}
E_1(x,z)+E_2(x,z)&=\frac{8\pi}{p}\Delta x\\
E_1(x,z)-E_2(x,z)&=\frac{4\pi}{\lambda}(\cos(\theta_{m\theta A})-\cos(\theta_{m\theta B}))\Delta z
\end{aligned}
\qquad [9.33]
$$

If the XZ encoder is designed as a transmissive or a reflective double-pass encoder (Holzapfel, 2009), then also the sensitivities would double in magnitude.

9.6 Current and future trends

For the high-end side, interferential encoder and interferometer systems will still be unsurpassed with respect to the attainable resolution and accuracy. For encoders, miniaturization and cost reduction is seen in the integration of encoder components (source, diffractive optics and detector) on a single chip (ChipEnc, 2004; Carr *et al.*, 2008). Miniaturization is also seen in fiber-based interferential encoders with integrated diffractive elements (Tobiason, 2005; Tobiason and Altendorf, 2009). For fiber-based interferometer and encoder systems, the light source may be shared among probes, and the signals of multiple axes may be multiplexed and read out sequentially by a single signal analyzer to reduce cost. Fiber-based short stroke interferometers may partly replace tactile systems for measuring surface structure or encoders for measuring stability (Lindner and Schmidke, 2009). As the cost of image sensors and image processing power has dropped significantly, another trend that can be observed is the development of vision-based image correlation and absolute encoders. An optical mouse is a good example of a low-cost vision-based image correlation sensor. In an optical mouse, a low resolution camera takes successive images of the surface on which the mouse operates. The surface is illuminated at grazing incidence by a light-emitting diode to enhance the contrast of the imaged texture. Alternatively, optical mice may use infrared lasers to generate a speckle image on the surface. Image correlation is used to detect the displacement with respect to a reference image using a dedicated processor. Based on the principle of an optical mouse, laser speckle correlation sensors are commercially available achieving sub-nanometer resolution (Jones *et al.*, 2001). In vision-based optical encoders, an image of a moving grid pattern is captured and analyzed for determining the absolute position of the object relative to the camera. The imaged patterns typically consist of high contrast marks or grid patterns. For absolute sensing, the grid typically contains a unique mark layout for every measurement position (Nilsagård *et al.*, 2009; Gribble and Robert, 2011). When imaging moving patterns, motion blur may be avoided using pulsed illumination or shuttered image acquisition while offering absolute timestamps for the recorded images. The measurement resolution may depend on the magnification of the optical system and the algorithms used to transform the observed images into a position or displacement signal. To reduce the out-of-plane sensitivity of a vision system, telecentric optics may be used. An example of a 1D absolute encoder (Gribble and Robert, 2011) contains a scale with a uniquely spaced barcode-like line pattern that is recorded by a high speed line camera and analyzed by a microprocessor to obtain the absolute scale position. In comparison with traditional encoder systems image recognition systems are more flexible in design and often can detect multiple degrees of freedom. The image processing may, however, require complex algorithms which may make it a challenge to design both accurate and fast algorithms to minimize signal delay.

9.7 Conclusion

The stability of laser interferometer systems depends on the stability of the wavelength which, in turn, is dependent on the laser frequency and the fluctuations of the refractive index of the medium (typically air or vacuum), or how well the wavelength variations can be calibrated inline using an (absolute) refractometer. The stability of an optical encoder mainly depends on the stability of the scale material (e.g. invar, glass, steel, zerodur) or how well the thermal expansion can be predicted using information from temperature sensors. Interferometers are traceable to SI standards by comparing the laser frequency with a reference laser, where scales are traceable to the SI standard by comparing the scale lines with a laser interferometer-based comparator which is ideally positioned in vacuum (Kunzman *et al.*, 1993; Maeda, 2001; Holzapfel, 2008). Linearity has two aspects: coarse errors and interpolation errors. Laser interferometers do not have coarse errors, while for scales the coarse errors must be calibrated. The differences can be reduced by application of calibrated scales and inline error correction. Interpolation errors of laser interferometers are comparable and typically in the range of 1–20 nm. For high-end encoders and interferometers, these periodic deviations can be reduced to sub-nanometer level by inline calibration. The measuring range is easily scalable to several meters for interferometers whereas for encoder systems the range is limited to the size at which the scales can be manufactured and mounted to a base with sufficient stability. In most machines with stapled linear actuators, it is easier to integrate a scale than an interferometer.

9.8 References

Benzaid, O. and Bird, B.M. (1993), 'Interpolation techniques for incremental encoders'. *Proceedings of the 23rd Int. Intelligent Motion Conf.*, 22–24 June, 165–172.

Birch, K.P. and Downs, M.J. (1993), 'An updated Edlén equation for the refractive index of air', *Metrologia*, **30**, 155–62.

Bobroff, N. (1987), 'Residual errors in laser interferometry from air turbulence and nonlinearity', *Applied Optics*, **26**(13), 2676–582.

Bönsch, G. and Potulski, E. (1998), 'Measurement of the refractive index of air and comparison with modified Edlén's formulae', *Metrologia*, **35**(2), 133–9.

Breyer, K.H., Pressel, H.G. and Zeiss, C. (1991), 'Principle of mounting a Zerodur scale. Paving the way to thermally stable coordinate measuring machines', in *Progress in Precision Engineering*, Springer-Verlag, Braunschweig, Germany, 56–76.

Carr., J., Desmulliez, M.P.Y., Weston, N., McKendrick, D., Cunningham, G., McFarland, G., Meredith, W., McKee, A., Langton, C. and Eddie, I. (Heriot-Watt University) (2008), 'Miniaturised optical encoder'. *Proceedings of the SPIE* Vol. 7068 70680M-1.

ChipEncoder™ (2004),' Integrated encoder (MicroE)', 23 February 2004, http://www.designnews.com/author.asp?section_id=1386&doc_id=223445.

Chu, D. and Ray, A. (2004), 'Nonlinearity measurement and correction of metrology data from an interferometer system', *Proceedings of the 4th Euspen International Conference*, 30 May–3 June, Glasgow, UK, 301–4.

Dürschmid, F. (1993), UntersuchungenzurMesssignalerzeugung und Messsignalvera rbeitunginkrementalerLangenmesssystemeunterbesondererBerucksichtigung der analyse und Reduzierungsystematischer und zufalligerMessabweichungen. PhD thesis.TechnischenUniversitätIllmenau, http://home.arcor.de/manfred.duerschmid/beruf/dissertation.pdf.

Edlén, B. (1966), 'The refractive index of air', *Metrologia,* **2**(2), 71–80.

Fan, K.-C. Bor-Cheng Lee and Yi-Cheng Chung (2008), 'A planar laser diffraction encoder in Littrow configuration for 2D nanometric positioning'. *Japanese Journal of Applied Physics*, **47**, 1833–7.

Gribble, A.P and Robert, I. (Renishaw) (2011), 'Absolute encoder setup indication'. *Patent No. US 2011/0173832 A1*, Issue date: 21 July 2011, related to the Renishaw RESOLUTE™ absolute encoder.

Hagiwara, N. (1992), 'A method of improving the resolution and accuracy of rotary encoders using a code compensation technique'. *IEEE Transactions on Instrumentation and Measurement*, **41**(1), 98–101.

Heydemann, P.L.M. (1981), 'Determination and correction of quadrature fringe measurement errors in interferometers', *Applied Optics*, **20**(19), 3382–4.

Holzapfel, W. and Linnemann, U. (Heidenhain GmbH) (2006). 'Position measuring arrangement', *Patent No. US7046368B2*, Issue date: 16 May 2006.

Holzapfel, W. (Dr. Johannes Heidenhain GmbH) (2008), 'Advancements in displacement metrology based on encoder systems', *ASPE Proceedings*, 19–24 October.

Holzapfel, W. (Heidenhain GmbH) (2009), 'Position measuring device', *Patent No. US 7573581 B2*, Issue date: 11 August 2009.

Huber, W., Holzapfel, W. and Kober, H.R. (Heidenhain) (1993), 'Interferential measurement device for at least one direction of measurement', *Patent No. US005264915A*, 23 November 1993.

Jones, B.K. and Nahum, M. (Mitutoyo Corp.) (2001), 'Systems and methods for reducing accumulated systematic errors in image correlation displacement sensing systems'. *Patent No. US7085431 B2*, 13 November 2001, related to 'Mitutoyo 'MICSYS' 2D correlation sensor.

Joo, K.N., Ellis, J.D., Spronck, J.W., van Kan, P.J.M. and Munnig Schmidt, R.H. (2009), 'Simple heterodyne laser interferometer with subnanometer periodic errors', *Optics Letters*, **34**(3), 386–8.

Joo, K.N., Ellis, J.D., Buice, E.S., Spronck, J.W. and Schmidt, RH. (2010), 'High resolution heterodyne interferometer without detectable periodic nonlinearity', *Optics Express*, **18**(2), 1159–65.

Kim, H.S., Schmitz, T.L., Beckwith, J.F. and Rueff, M.C. (2010), 'A new heterodyne interferometer with zero periodic error and tunable beat frequency', *Journal of Optics Research*, **11**(3), 213–244.

Kniel, K. and Schwenke, H. (2007): Genauigkeitssteigerung und KalibrierungvonKoordinatenmessgerätenmitdemLaserTRACER, In: PTB-Mitteilungen, *ThemenschwerpunktKoordinatenmesstechnik*, Sonderdruck, ISSN 0030-834X.

Kuhn, R. (Heidenhain) (2010),'1Dplus encoder', *Encoders for the Nanotech World*, Traunreut, Germany, 10 September 2010. http://www.designworldonline.com/articles/6391/305/Encoders-for-the-Nanotech-World.aspx.

Kunzman, H., Pfeifer T. and Flugge, J. (1993), 'Scales vs. laser interferometers, performance and comparison of two measuring systems', *Annals of the CIRP*, **42**(2), 753–67.

Lindner, M. and Schmidke, B. (Bosch GmbH) (2009), 'Interferometric measuring device for recording geometric data for surfaces', *Patent No. US7576864B2*,Issue date: 18 August 2009.

Maeda, F. (Mitutoyo corp.) (2001), 'A line scale measurement system, vacuum laser interferometer mounted 1.6 m ultra-high precision length measuring machine', *Lectured at 159.PTB Seminar in November 2001*.

Mitutoyo Catalog No.E4189-539, 579 (2012). Mitutoyo Absolute glass transparent linear encoder AT332, AT353, AT553.page 11. Available at http://www.mitutoyo.com/pdf/E4189-539_579NC_Scale.pdf

Nilsagård, J., Stenberg, M. and Takman, O. (Gcoder systems AB) (2009), 'Positioning pattern'. *US 2009/0177416 A1*, 9 July 2009.

Patorski, K. (1986), 'Incoherent superimposition of multiple self-imaging under plane wavefront illumination'. *Applied Optics*, **25**(14), 2396–403.

Peggs, G.N. (1999), 'Advanced CMM design', Invited paper, National Physical Laboratory, *1999,* Volume 20, *ASPE Proceedings*.

Quenelle, R.C. (1983), 'Nonlinearity in interferometer measurements', *Hewlett-Packard Journal*, **34**(4), 10.

Ruijl, T.A.M. (2001), 'Ultra precision coordinate measuring machine, design calibration and error compensation', *PhD thesis*, Technische Universiteit Delft, Philips Centre for Industrial Technology Eindhoven, The Netherlands, ISBN 90-6464-287-7.

Sandoz, P. (2005), 'Nanometric position and displacement measurement of the six degrees of freedom by means of a patterned surface element', *Applied Optics*, **44**(8), 1449–53.

Schellekens, P., Wilkening, G., Reinboth, F., Downs, M.J., Birch, K.P. and Spronck, J. (1986), 'Measurement of the refractive index of air using interference refractometers', *Metrologia,* **22**, 279–87.

Tobiason, J. (Mitutoyo Corp.) (2005), 'High accuracy miniature grating encoder readhead using fiber optic receiver channels'. *Patent No. US6906315B2*, Issue date: 15 June 2005.

Tobiason, J. and Altendorf, H. (Mitutoyo Corp.) (2009), 'Fiber-optic miniature encoder for fine pitch scales'. *Patent No. US7502122B2*, Issue date: 10 March 2009.

Van Seggelen, J.K. (2007), 'Design of a 3D-coordinate measuring machine for measuring small products in array setups', *PhD thesis*, Technische Universiteit Eindhoven, The Netherlands, ISBN 978-90-386-2629-1.

Vermeulen, M.M.P.A. (1999), 'High precision 3D-coordinate measuring machine, design and prototype-development', *PhD thesis*, Technische Universiteit Eindhoven, The Netherlands, ISBN 90-386-2631-2.

Wu, C. Lawall, J. and Deslattes, R.D (1999), 'Heterodyne interferometer with subatomic periodic nonlinearity', *Applied Optics*, **38**(19), 4089–4094.

10

Microfabrication technologies used for creating smart devices for industrial applications

J. M. QUERO, F. PERDIGONES and
C. ARACIL, University of Sevilla, Spain

DOI: 10.1533/9780857099297.2.281

Abstract: The purpose of this chapter is to provide the reader with a general overview of microsystems technology. It introduces the design and process flow in the microfabrication of a device: the materials used in the microfabrication, the essential microfabrication processes and the simulation tools involved in the design towards smart devices.

Key words: microsystems, micro-electro-mechanical systems (MEMS), microfabrication, silicon, polymers, photoresist, photolithography, deposition, etching, bonding, simulation.

10.1 Introduction

Micro-electro-mechanical systems (MEMS) sensors are becoming a fundamental building block in the realization of smart sensing systems, as they provide critical added value such as miniaturization, low cost and low power consumption, which are needed in today's applications.

Microsystems is a discipline that produces systems at a micro scale. It studies materials and processes for the realization of very small devices and systems, with similar features to existing macro systems, but with quite different properties and behavior. The development of these devices, such as accelerometers or pressure sensors, is not its only goal. Furthermore, the microfabrication processes should be regarded as manufacturing capabilities that could be used in a synergistic way with standard processes in industry to produce better or even new products.

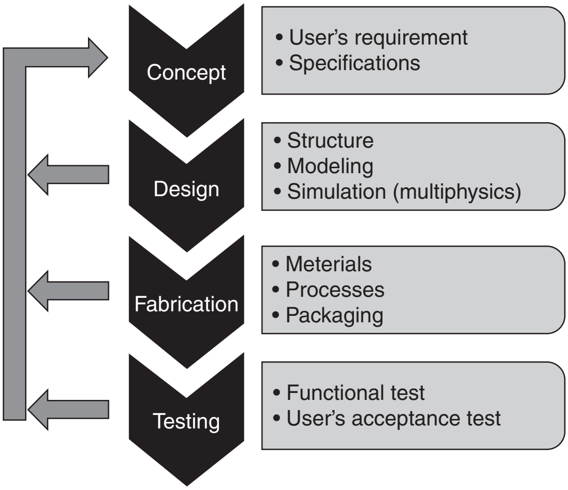
Microsystems technology is as old as the integrated electronic circuit technology that was developed in the 1950s. One of the key steps in the microelectronics fabrication process is the metal removal using chemical etching. This process is a precursor of the metal sacrificial process used in surface micromachining. However, the deployment of the metal sacrificial process has been much slower than the metal removal process in integrated electronics. This is for two reasons: the difficulties in guaranteeing complex but reliable processes and the fact that the great success of microelectronics has eclipsed microsystems. However, nowadays there is a set of well-defined processes that are sufficiently mature to be used for mass production.

The purpose of this chapter is to provide the reader with a practical overview of microsystems technology, including all aspects to be considered in the manufacturing process flow of the microfabrication of a device: the MEMS design and modeling; the selection of materials used in the microfabrication, such as those for structural and sacrificial use; a description of the most common microfabrication processes, related to silicon and polymers; and the simulation tools involved in its design prior to its fabrication, which improves its reliability. This design benefits from microtechnology as it guarantees strong performance and compactness in the final products.

10.2 MEMS design and modeling

The manufacturing process of a microsystems device consists of a set of phases that are common to general industry. These phases include the development of the concept, its design and simulation, the selection of materials and manufacturing processes and testing, as depicted in Fig. 10.1.

After finalizing any of these phases, the results are checked against the initial specifications and, usually, iterations are needed to fulfill the objectives.


Concept
• User's requirement
• Specifications
Design
• Structure
• Modeling
• Simulation (multiphysics)
Fabrication
• Meterials
• Processes
• Packaging
Testing
• Functional test
• User's acceptance test

10.1 Design flow of the manufacturing process of a microsystem.

Due to the fact that the minimum size of the device is typically above 1 μm, the continuum hypothesis is valid and the mechanical, thermal and electrical laws of the macroscopic world are applicable in the modeling of the device. This fact also allows for the use of conventional design tools, provided they consider the scaling laws that occur at this level. For example, when studying the behavior of the fluids at micron size, Navier–Stokes equations are valid. However, phenomena that are usually neglected, like surface tension, now contribute significantly. This is why specific computer aided design (CAD) software is utilized during the simulation phase.

The substance that is used as the base material is called the 'substrate'. The more common substrates are silicon, glass and polymers in the shape of wafers. When working at micron size, the designer has the advantage of using very small quantities of materials, so a large variety of expensive materials can be used to produce a structure, like gold or platinum.

Another important aspect to be considered is the selection of the manufacturing processes. The definition of a run card (i.e. a table containing the full sequence of processes and their parameters required to fabricate a MEMS device) is a key issue for the industrial success of a microsystem. A large variety of MEMS processes are available in a clean room (as will be described in Section 10.4), even for a specific procedure (e.g. the physical deposition of material). There are several criteria that should be applied in this selection process: the most important is the compatibility of the process with previous processes, because an aggressive process can negatively alter, or even destroy, an intermediate structure. It is also important to choose a procedure that achieves the required performance at a low cost. Finally, a designer should consider other more conventional aspects like yield and reliability. A practical constraint that is imposed when selecting the processes is their availability in a given foundry, because it is desirable to run all of the processes in a single clean room so as to avoid contamination and other side effects.

10.3 Materials

The design of the device must allow for the restrictions that material and process flow impose. Therefore, knowledge of microfabrication processes and materials is important when selecting those most suitable for the device that is being designed. The variety of materials and processes is so great that a method for making these decisions has been developed in order to establish the optimal choices (Quinn *et al.*, 2006). However, this chapter focuses on the most common basic processes and some widely used materials, starting with the materials that are most often used in MEMS: silicon, photoresists, polymers and metals.

10.3.1 Silicon

Traditionally, MEMS have been linked to silicon and silicon-based materials, due to its origin in microelectronics. The use of other materials – such as silicon dioxide, silicon nitride, silicon on insulators (SOI), gallium arsenide, quartz, glass and diamond – has also been explored. However, silicon, polysilicon and amorphous silicon are the most common materials currently used in MEMS for commercial production.

Silicon is identified by a series of valuable physical properties that have played a fundamental role in micromachining (Dargys and Kundrotas, 1994). Moreover, it is a well-characterized material, due to its use in the microelectronics industry. One of the advantages from the microfabrication point of view is that it has different crystallographic structures, which affects the electrical, mechanical and optical properties of the material. There are different forms of silicon:

- *Amorphous*: This is characterized by an unpredictable long-range atomic order, and there is no well-defined band gap.
- *Polycrystalline*: This presents a long-range order. The solid is composed of many small crystal blocks located together. There is a distinct band gap.
- *Crystalline*: In this case, there is an extremely long-term order, with little difference in the placement of the atoms throughout the solid. The band gap is therefore well-defined. Material properties of the crystal are highly reproducible in this case, but they will generally depend on the direction within the crystal, so the material is said to be 'anisotropic'. In a crystal, it is possible to identify the position of the atoms using a system of coordinates whose axes are placed along the edge of the lattice. Miller planes are represented by (x, y, z) where the (x, y, z) planes are perpendicular to the corresponding vectors (Lima de Faria, 1990). Through Miller indices, the different anisotropic etching rates are known, and it is possible to create structural forms with different aspect ratios.

10.3.2 Polymers

Polymers are macromolecules, usually organic ones, formed by the joining of smaller molecules called 'monomers'. Polymerization is the process by which a polymer synthesizes from monomers. Its application in the MEMS industry has led to the research and development of new polymers which has had a positive effect on the polymer industry.

Polymers differ in their mechanical properties; optical properties; temperature stability; resistance to chemicals such as acids, alkalis or organic solutions; and there are also biodegradable polymers. There is a suitable polymer for each application.

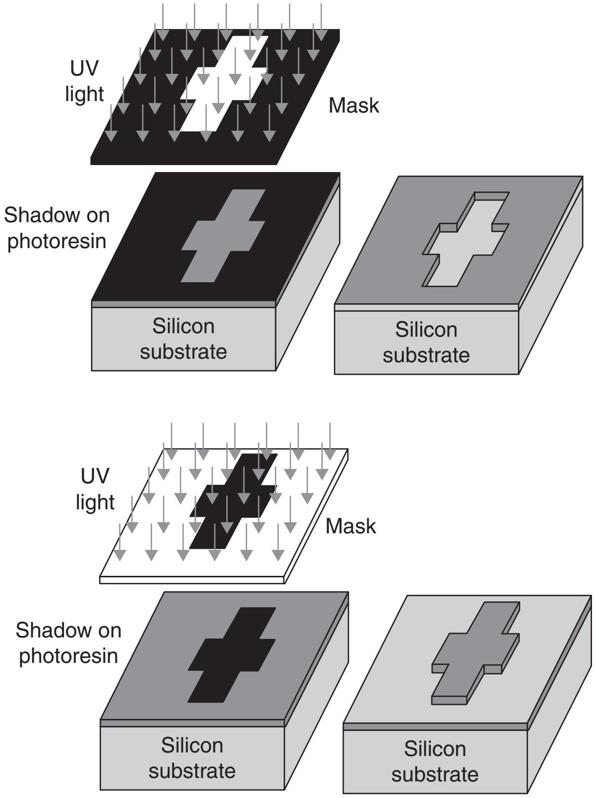
A possible classification of polymers is based on the type of monomers of which they are composed. In this way, homopolymers are composed of the same type of monomers, while copolymers are made from different types of monomers. The copolymers are developed to modify the physical, mechanical, thermal and electrical properties of manufactured plastics in order to meet specific needs (Hadjichristidis *et al.*, 2003).

Two polymers that are commonly used in MEMS are polymethylmetacrilate (PMMA) (Heckele and Schomburg, 2004), and polydimethylsiloxane (PDMS) (Mata *et al.*, 2005). Both of these are well-known, due to the advantages in their fabrication, as well as their internal characteristics. There are also polymers, like SU-8, that cure depending on a previous exposition to a specific wavelength radiation. The photoresists belong to this photosensitive polymer family. Due to their relevance in photolithography, a complete section has been devoted to them.

10.3.3 Photoresists

Photoresists are fundamental materials related to photolithography. They are light-sensitive materials, composed of a polymer, a sensitizer and a solvent. Each element has a particular function. The polymer changes its structure when it is exposed to radiation. The solvent allows the photoresist to be spun and to form thin layers over the wafer surface. Finally, the sensitizer, or inhibitor, controls the photochemical reaction in the polymer phase.

Photoresists can be classified as positive or negative. In the positive photoresists, the photochemical reaction that occurs during exposure weakens the polymer, making it more soluble to the developer, so the positive pattern is achieved. The mask, therefore, contains an exact copy of the pattern which is to remain on the wafer, as a stencil for subsequent processing. In the case of negative photoresists, exposure to light causes the polymerization of the photoresist, and therefore the negative resist remains on the surface of the substrate where it is exposed, and the developer solution removes only the unexposed areas. Masks used for negative photoresists, therefore, contain the inverse or photographic ‘negative’ of the pattern to be transferred. Both negative and positive photoresists have advantages and disadvantages. The advantages of negative photoresists are good adhesion to silicon, lower cost and a shorter processing time. The advantages of positive photoresists are better resolution and thermal stability. Figure 10.2 shows the result of the exposure of both photoresists. As can be seen, the same mask produces complementary structures.


UV light
Mask
Shadow on photoresin
Silicon substrate
Silicon substrate
UV light
Mask
Shadow on photoresin
Silicon substrate
Silicon substrate

10.2 Negative and positive photoresists are exposed throughout a mask. The result of both cases is complementary.

In order to improve the adhesion of the photoresist, a prior preparation of the substrate is usually carried out. A promoter, which is a material that gives rise to strong adhesion of the photoresist to the substrate, is applied. Some promoters are deposited by spinning, while others are deposited by introducing the substrate to a desiccator with the promoter, so that it can be adhered by evaporation.

10.3.4 Metals

Other materials that have been explored in MEMS are metals. They do not have the same properties as silicon but, when used within their limitations, metals can exhibit high levels of reliability. Many microfabrication processes are related to deposition of metals (e.g. electroplating, evaporation and sputtering processes). Commonly used metals include gold, nickel, aluminium, copper, chromium, titanium, tungsten, platinum and silver.

10.3.5 Piezoelectric and piezoresistive materials

An important collection of MEMS sensors are based on the piezoelectric or piezoresistive effect, which generates an electrical moment or a change of resistivity by changing the stress that is applied to a solid. With the piezoelectric effect, most piezoelectric transducers are manufactured from ceramic materials, usually from the lead zirconate titanate (PZT) family, due to their excellent piezoelectric parameters, thermal stability and dielectric properties. Recently, the discovery of piezoelectricity in polymeric materials has driven the development of innovative devices (Tadigadapa and Mateti, 2009).

The processes related to these materials cover a broad range including deposition of piezoelectric thin films, sputtering or chemical vapor deposition (CVD) techniques and synthesizing by conventional methods. The method chosen depends on the kind of materials that are to be used.

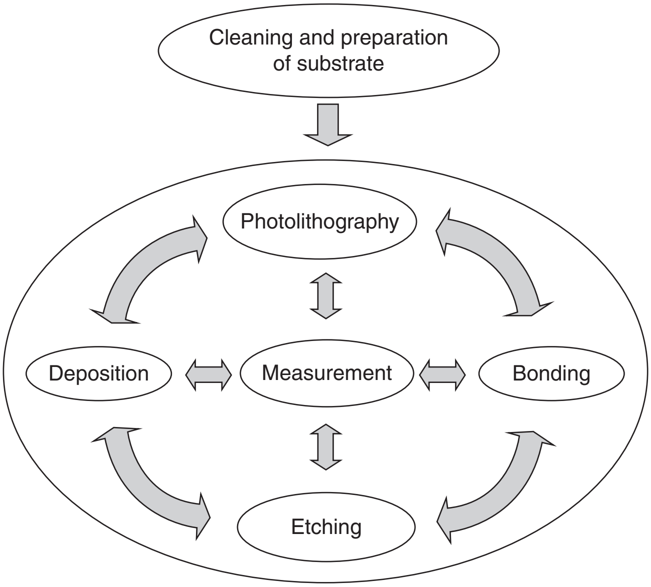
In the case of the piezoresistive effect, semiconductors are characterized by this effect, so materials such as germanium, polycrystalline silicon, amorphous silicon and single crystal silicon are used. Thus, the microfabrication of these devices is related to integrated circuit (IC) technology (Barlian *et al.*, 2009).

10.4 Microfabrication processes

Manufacturing processes have evolved in recent decades. Significant research lines have been associated with the search for new manufacturing processes for MEMS. Since the range of processes is large and cannot be covered in a book of this nature, only the basic processes are explained. For a deeper study, Madou (2002) is recommended.

It should be remembered that significant parts of the technology have been inherited from IC technology. Some shared factors are the use of silicon wafers as the most common substrate, the fact that the structures are realized in thin films of materials, the transfer of a pattern using photolithographic methods and so on. Regarding infrastructure, microfabrication is carried out in a specialized facility, called a ‘clean room’. This is a room that has a controlled environment, where the concentration of airborne particles is limited to specified values.

One of the classifications of microfabrication technologies divides it into two main categories: silicon and soft technologies, the latter including the use of plastics or elastomers. Silicon technology is also often classified in a hard technologies category, since it makes use of hard materials, such as silicon or glass. However, this distinction is not entirely clear-cut and, in some



10.3 Schematic of a generic manufacturing process flow.

cases, silicon is still necessary in the process flow, but not in the final device. With regard to the maturity of these technologies, the scenario is completely different. While silicon technology is standardized, the techniques associated with polymers and plastics are still being developed, and new processes and even new materials are continuously being created (Buddhadev *et al.*, 2010; Gogolides *et al.*, 2010).

In silicon technology, there are two main groups: additive techniques and etching techniques (MEMS and Nanotechnology Clearinghouse: available from http://www.mems-exchange.org/MEMS/processes (2010)). The general flow process is shown in Fig. 10.3. This flow chart is composed of basic cyclic processes that are repeated as many times as necessary to perform the different layers of the design.

First, a typical fabrication process starts with a wafer (silicon, polymer, glass, etc.) that may play an active role in the final device or may only be a substrate on which the MEMS is built. Preparation of the wafers is required, usually by an RCA process (this acronym comes from Radio Corporation of America, the company that designed this process). Then, there are three basic building blocks in MEMS technology: the deposition of thin films of material on a substrate, the application of a patterned mask on top of the films by photolithographic imaging, and the selective etching of the films onto the mask. These processes are repeated until the device is finished. These are explained in the following subsections.

10.4.1 Photolithography

Photolithography is a fundamental process in microfabrication. It is a combination of photo-patterning and etching. The key material is the photoresist.

First, the photoresist is spin-coated over the substrate; this process is also called 'spin casting'. During this process, a small amount of the photoresist is injected by a nozzle onto a rotating disk. The disk is then accelerated to a higher speed to sweep off the photoresist and the wafer surface is therefore completely covered. The deposition thickness depends on the spin speed, the solution concentration and its molecular weight (measured by intrinsic viscosity).

The photoresist must be baked in order to eliminate the solvent. Once the deposition is completed by the spin coater, a homogeneous layer of photoresist is obtained and can be exposed to UV radiation. The purpose of the exposure is to provide light intensity, direction and uniformity of the characteristic spectrum that allows the transfer of the mask in the resin.

A previous alignment has to be achieved with regard to the mask that will be transferred. The mask usually consists of a wafer glass with a patterned metal emulsion on one side, the shape of which will be transferred. To obtain a lower-resolution mask, acetate sheets that have already been plotted are used. The cost and resolution related to each mask is completely different.

The machine at the center of the whole exposure process is called a 'mask-aligner'. Two main elements are included: a collimated light source, and an optic system to align the mask. It also includes a filter to increase the resolution obtained on the walls, which removes all the wavelengths except UV, the wavelength to which the photoresist is sensitive. The mask is properly aligned within the substrate by including alignment marks. Once aligned, the photoresist is exposed to light through the mask. There are three types of exposure: contact, proximity or projection (Maluf, 2002). In the first case, the mask is in direct contact with the photoresist that will be exposed. Due to the proximity of the two elements, high precision can be gained; however, some wastes can be blocked between the mask and the wafer, which can damage the mask and cause defects. The second option is very similar to the first, but there is a distance of about 10–25 μm between the mask and the wafer. This separation decreases the defects, but they are not completely eliminated. Finally, the projection system holds a separation of centimeters between the mask and the wafer, avoiding the possibility of defects in the mask.

Technical datasheets of resins include information about general exposure requirements, but the optimal time of the exposure dose must be chosen by taking into account the type of light source, its power and spectral range and so on.

The development process is the final patterning step. It is responsible for the transformation of the latent image on the resin during exposure to an

Table 10.1 Deposition techniques

Additive technique	
Deposition	
Chemical reaction involved	Chemical vapor deposition (CVD), Low pressure CVD (LPCVD) Plasma enhanced CVD (PECVD) Electrodeposition or electroplating Epitaxy Thermal oxidation
Physical vapor depostion (PVD)	Evaporation Sputtering
Spin casting	

UV image. The selective dissolution of the photoresist is carried out during this process and, according to the type of photoresist that is used (i.e. the illuminated area or the dark), it will be soluble.

There are two types of techniques: wet and dry development. The most commonly used is the wet technique, using spray or immersion. Dry developing is used in some high-resolution applications. The development process can be modeled as an isotropic etching process, which will be studied in a subsequent section, although certain parameters vary.

The choice of developer differs according to the type of photoresist. With positive photoresists, the developers are usually alkaline aqueous solutions, whereas in the case of negative photoresists, negative organic developers are usually used.

Finally, after the wafer has been developed, it is necessary to wash and dry the wafer to remove any developer as this could cause damage.

10.4.2 Deposition

The deposition of materials can be achieved through a variety of techniques, which are continuously evolving. The MEMS deposition technology in Table 10.1 can be classified into two groups based on whether the nature of the reaction chemical or physical.

Chemical-deposition processes

There are different techniques in which a chemical reaction is involved:

- *Chemical vapor deposition (CVD)*: This is achieved using a chemical-deposition reactor, into which the substrate is placed. A number of gases are added into the reactor and are responsible for the chemical reaction that takes place. The product of the reaction is a solid material that condenses on the surfaces inside the reactor, where the wafer is located.

Low pressure CVD (LPCVD) and plasma enhanced CVD (PECVD) are the most important CVD technologies in MEMS. Their characteristics are quite different. The LPCVD process produces layers with exceptional uniformity of thickness and material characteristics; the quality of the films of PECVD tend to be inferior due to a lower working temperature. Moreover, LPCVD systems deposit films on both sides and over many wafers at a time, unlike PECVD. However, the deposition rate for LPCVD is slower.

- *Electrodeposition or electroplating*: This is limited to electrically conductive materials. There are essentially two technologies: electroplating and electroless plating.

  In the electroplating process, the substrate is immersed in a liquid solution called an 'electrolyte'. Then, an electrical potential is applied between a conductive area on the substrate and an electrode. A resulting chemical redox process generates a layer of material at the counter-electrode.

  In the case of the electroless plating process, the deposition takes place over any surface that forms a sufficiently high electrochemical potential with the solution. This method is advantageous since it does not require any external electrical potential or contact with the substrate during processing. Unfortunately, the chemical reaction that takes place is more complex than in the case of electroplating, and it is more difficult to control the film thickness and uniformity.
- *Epitaxy*: This technology makes possible to build, over a semiconductor crystal that acts as a seed, a new deposition with the same crystallographic orientation. The growth is achieved in a vapor-phase chemical-deposition reactor, into which source gases are added. The process involves dissociation or hydrogen reductions at a high temperature (>800°C). An advantage of epitaxy is the high growth rate of the material and its considerable thickness (>100 μm).
- *Thermal oxidation*: This process involves the oxidation of the substrate surface within an oxygen rich atmosphere. Thus, this process is limited to materials that can be oxidized, and it can only generate films that are oxides of that material. This technology consumes some of the substrate because the growth of the film is achieved by diffusing oxygen into the substrate. The rate of oxide growth is often predicted using the Deal–Grove model (Deal and Grove, 1965). Usually, this process is used to create films of silicon dioxide.

Physical-deposition processes

In physical-deposition processes, the supplied material is physically deposited onto the substrate. Some of the physical vapor deposition (PVD) techniques are:

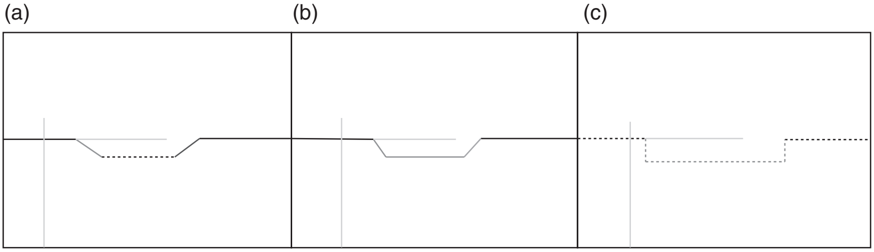
- *Evaporation*: In this technology, the substrate and a source of the material to be deposited are located inside a vacuum chamber. The vacuum is required in order to allow the molecules of source material to evaporate freely in the chamber, and therefore to be deposited on all surfaces. The source material is then heated until it boils and evaporates. Two technologies are based on this principle, but the method used to heat the source material differs. These techniques are e-beam evaporation and resistive evaporation. According to the material, an evaporation method can be used that is related to the phase transition properties of that material.
- *Sputtering*: In this case, the substrate and the source material (called the 'target') are also placed into a vacuum chamber where an inert gas is introduced at low pressure. Gas plasma is achieved using a radio frequency (RF) power source, causing the gas to become ionized. The target is bombarded by the ions, causing atoms of the source material to be ejected and then deposited over the substrate. This technique works at a lower temperature than evaporation.
- *Spin casting*: This process has already been explained in the photolithography section, but there is additional information. In spin casting, the material is deposited by spraying or spinning, so the material to be deposited has to be dissolved in a solvent to gain a liquid state. After a baking process, once the solvent has evaporated, a thin film of the material remains on the substrate. This process is particularly useful for the production of polymer layers. The range of thicknesses that can be obtained by this technique ranges from a single monolayer of molecules to micrometers.

10.4.3 Etching

A process in the flow chart (Fig. 10.3) that is necessary to achieve a MEMS structure is the etching of previously deposited thin films and/or the substrate itself. In general, there are two categories of etching processes (see Table 10.2), wet etching or dry etching. The method used depends upon whether the material is immersed in a chemical solution or sputtered or dissolved using reactive ions or a vapor-phase etchant.

Table 10.2 Etching techniques

Wet etching Dry etching	Reactive ion etching (RIE) Vapor-phase etching Sputter etching	Deep reactive ion etching (DRIE)



10.4 Simulations of etching processes using ACES (Anisotropic Crystalline Etch Simulator) for different wafers. The cross sections present different orientations of the crystallographic structure, so the etching shapes differ. In case (a) face ⟨110⟩ and edge ⟨100⟩, case (b) face ⟨100⟩ and edge ⟨110⟩ and (c) face ⟨100⟩ and edge ⟨100⟩.

Wet etching

This consists of immersing the wafer in a container with a liquid solution that will dissolve the material. On one hand, it is the simplest etching technique, and simple and low-cost facilities are required. On the other hand, the etching process is very limited in many aspects. For example, the selection of the mask is complex because one must be very selective when choosing the materials: it must not dissolve, or it must at least etch more slowly than the material to be patterned. Moreover, many materials exhibit anisotropic etching in certain chemicals, so different etching rates in different directions are produced, as can be seen in Fig. 10.4. The images correspond to simulations of etching processes achieved by anisotropic crystalline etch simulators (ACES). This kind of etching allows for the creation of structures as micro-channels, micro-nozzles, micro chambers and so on.

Dry etching

Three separate technologies are included in this category: reactive ion etching (RIE), sputter etching and vapor phase. They generally consist of an advanced etching process that yields patterns with a high resolution.

- *Reactive ion etching (RIE)*: In this process, plasma drives a chemical reaction to remove materials. There is also a physical process similar to that of sputtering deposition, but which results in an etching. It is a complex task because chemical and physical etchings have to be balanced, so there are many parameters to be adjusted.

  A particular version of RIE is deep RIE (DRIE). In this process, etch depths of hundreds of microns can be achieved with almost vertical sidewalls and therefore high aspect ratio structures can be achieved. The technology is based on the process patented by Bosch (Laermer and Schilp, 2006). Two different gas compositions are involved. The first creates a polymer on the surface of the substrate, and the second etches the substrate. Since the polymer protects the sidewalls from etching, etching aspect ratios of 50:1 can be achieved.

- *Sputter etching*: This is very similar to RIE, but there are no reactive ions and it is based on a sputtering process where the substrate is subject to ion bombardment, instead of the material target.
- *Vapor-phase etching*: This is a kind of isotropic etching, which is a similar process to RIE, although it requires a simpler setup. The wafer is also placed inside a chamber into which one or more gases are introduced. The target material is superficially dissolved due to a chemical reaction with the gas molecules. The two most common reactive vapor-phase etching technologies are silicon etching using xenon diflouride (XeF_2) and silicon dioxide etching using hydrogen fluoride (HF). Suspended and 3D structures can be obtained.

10.4.4 Bonding techniques

A basic step in the fabrication of industrial sensors and actuators is the bonding of the wafers. Typical MEMS devices such as accelerometers or pressure sensors require a cavity to house the micromachined free-standing structures that comprise them (Henmi *et al.*, 1994; Schmidt, 1998; Candler *et al.*, 2003; Tekin, 2011). Both the cavity and the microstructure are fabricated in a separated wafer, and then they are bonded using the common technique of wafer-to-wafer bonding. The bonding can be performed in a vacuum ambient in order to reduce the damping effect in microdevices, or to create reference pressure chambers. Also, the fabrication of multilevel chips (Miki *et al.*, 2003) or structures requires the use of bonding.

There are two substrates that are mainly used in MEMS: silicon and glass. Typically, the structures are fabricated using silicon, and the cavities material can be silicon or glass.

The most common techniques used in MEMS to perform the silicon-to-silicon, glass-to-silicon bonding are presented in Table 10.3. They are fusion bonding, anodic bonding and eutectic bonding. Also, there are other techniques to bond the wafers, such as the use of adhesive materials; for example, SU-8. Apart from these, there are other techniques to bond different MEMS materials, essentially polymers, for instance SU-8-to-SU-8, PDMS-to-PDMS, PDMS-to-SU-8, even Silicon-to-PDMS or Glass-to-PDMS.

Nowadays, silicon is used more than polymers in industrial applications. However, polymers are widely used for microfluidics applications, but mainly in rapid prototyping rather than mass production. Therefore, in the following sections, the three main techniques used for industrial applications are described.

Table 10.3 Bonding techniques

Fusion bonding
Anodic bonding
Eutectic bonding

Fusion bonding

This technique, also known as 'direct bonding' (Shimbo *et al*., 1986; Barth, 1990; Plößl and Kräuter, 1999), is widely used for silicon-to-silicon bonding. However, silicon-to-glass bonding can be also attained (Xiao *et al*., 1999) with this technique. Silicon fusion bonding (SFB) is based on the tendency that flat and smooth surfaces have to stick to each other. An annealing process is performed in order to complete the bonding. In general, fusion bonding has three steps: preprocessing (cleaning and treatment of the surface substrates), prebonding (the contact between surfaces) and annealing.

The process starts with the precleaning of the silicon wafers using, for example, a mixture of H_2O_2 and H_2SO_4 (piranha solution) and then dipping it in a dilute HF solution. Then, the wafers are cleaned following the industrial standard procedure SC (Standard Clean) purification devised by RCA. It involves two steps, RCA 1 and RCA 2.

$$\text{RCA 1}\left(\text{NH}_4\text{OH }(29\%)+\text{H}_2\text{O}_2\,(30\%)+\text{Deionized}-\text{H}_2\text{O }[1:1:5]\right)$$

$$\text{RCA 2}\left(\text{HCl }(37\%)+\text{H}_2\text{O}_2\,(30\%)+\text{Deionized}-\text{H}_2\text{O }[1:1:6]\right)$$

At the end of this step, the wafers are ready to be pre-bonded. In order to do so, the surface wafers are brought into contact. In this step, the wafers stick due to the flatness, smoothness and the hydrated nature of the surfaces. Then, the wafers are annealed in nitrogen ambient in an oven (at approximately 1100°C) for 2 h.

Anodic bonding

This technique, also called 'field assisted bonding', is mainly used for silicon-to-glass bonding (Henmi *et al*., 1994; Lee *et al*., 2000). However, silicon-to-silicon or glass-to-glass bonding can also be performed (Hanneborg *et al*., 1991; Kutchoukov *et al*., 2004), although it is less common. The anodic bonding is based on the use of electrostatic fields and high temperatures. The wafers must be flat and smooth, and have a similar coefficient of thermal expansion (CTE).

The process has two steps: cleaning, and application of high voltage and heat. The surfaces of both the silicon and glass wafers are cleaned in RCA solutions at a high temperature, in the same way as for SFB. Then, the surfaces are brought into contact on a hot plate at a temperature of 300–500°C and a voltage of 200–1000 V is applied. In order to achieve the bonding, the glass wafer is used as the cathode, and the silicon wafer is the anode. Finally, the bonded wafers are cooled down carefully to avoid causing thermal shock.

Eutectic bonding

Eutectic bonding (Wolffenbuttel, 1994, 1997) is one of the intermediate layer bonding techniques, together with solder, polymer, low melting-temperature glasses, and thermocompression, among others. This method is used to bond silicon-to-silicon using an intermediate metal layer; for example, gold or aluminum. The main idea of eutectic bonding consists of achieving a low melting-point alloy in order to bond the substrates by fusion. This technique is similar to fusion bonding but the annealing is performed at a much lower temperature.

The process has three steps: cleaning, metal deposition and pressure-temperature treatment. The cleaning of the wafer removes the natural oxide layer on silicon wafers in order to achieve a good adherence between the metal and the silicon. In order to do this, the wafers are dipped in a dilute HF solution. There are other methods to improve the adherence; for example, the use of auxiliary intermediate layers. Then, in the metal deposition step, a gold layer of 1000–2000 Å is sputtered over the wafers, or deposited by evaporation. Finally, the pressure-temperature step is performed in a vacuum ambient using a program of ramped temperatures and pressures. Gold-silicon eutectic formation occurs at 363°C, with 19% Si when the silicon diffuses into the gold. Thus, the temperatures used in this step are higher than 363°C, typically in the range of 410–450°C. First, a pressure of 2000–5000 mbar is applied, and then the heat is ramped up to achieve the desired temperature. This temperature remains constant until the bonding is achieved, due to the liquid state of the alloy. Finally, the wafers are cooled down using a temperature ramp.

This is a summary of the more common and more important bonding techniques used for sensors and actuators in industrial applications. Each method has advantages and drawbacks which are briefly commented upon below.

SFB provides high bonding strength and high stability to temperature since the final step of the process is performed at high temperature. However, this method requires advanced equipment and the surfaces must fulfill demanding specifications. Also, the electrical interconnects between parts of the device present problems in different wafers. Unlike SFB, the anodic bonding process is easy to perform, since bonding is at a low temperature. Also, the surfaces must fulfill the same requirements that are demanded by SFB, but in a more flexible manner. In addition, this technique provides transparent shells. However, since silicon and glass have different CTEs, it is possible that the wafers will deform. Eutectic bonding enables low temperature bonding, using inexpensive equipment. The surfaces must be smooth, but there is a greater tolerance in this matter than in other methods. However, eutectic bonding also presents a different CTE between materials. Furthermore, the flow during the annealing step cannot be homogeneous, leading to nonhomogeneous bonding.

10.5 Simulation

Simulations are an important issue in the fabrication of a device in any industrial field, in particular in MEMS. This consists of using a computer to study the device's behavior before fabrication, taking into account equations that are very difficult or impossible to solve manually in a reasonable time. Those equations involve second-order terms for couples problems (e.g. mechanical, electrical, thermal, magnetic and optical).

Simulation enables the avoidance of many problems during the development of the device, from designing to testing. The simulation helps during the design stage because it allows certain device parameters to be changed and the behavior as a function of those parameters' variations to be studied. Therefore, they reduce the risk of failure when testing the fabricated device. This matter is very important because it avoids the unnecessary cost of additional fabrication processes.

Important issues must be taken into account when performing a simulation. For example, it is important to consider the proposed model of the whole problem, the properties of the materials, the convergence study and the application of boundary and initial conditions. All these points will be commented on later.

There are several simulation programs for microsystems technologies, for example, COMSOL, ANSYS and CoventorWare. An example of simulations using CoventorWare will be presented in order to explain the next sections in a practical way.

10.5.1 Simulations and fabrication materials

The choice of effective material properties contributes to achieving simulations with final results that are in agreement with experimental results.

One material can have different properties depending on the process parameters that produce it (e.g. annealing times, percentage of certain additives and so on).

The software that is used to present the advantages of making simulations (CoventorWare) includes the main materials for MEMS applications. The properties of the material can be changed if necessary, and materials can be created by the designer. The simulations can include one or more materials, even solid and fluid materials interacting in the same simulation.

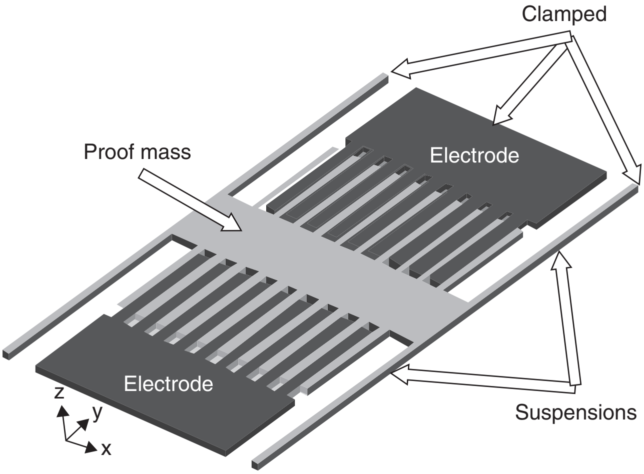
The example that follows consists of a capacitive acceleration sensor that could be applied in the automotive industry. The material of the structure that composes the sensor is polysilicon.

10.5.2 Solid model

The solid model is a 3D replica of the device to be simulated and built by software using the design dimensions and the materials properties. This replica can be obtained in various ways; for example by using the simulator, or by using auxiliary software that imports the model to the simulator. The software that is used to perform the simulation includes the option of making the solid model. In order to do so, the fabrication process of the device and the photolithographic mask must be defined. Once the masks have been drawn and the process configured, the simulator builds the structure. The result of this step is an exact model of the device. However, all fabrication processes have tolerances. These tolerances have an influence on the deviation of simulation results regarding final experiments in testing. The tolerances are not the only cause of deviation; there can also be deviation if the second- or third-order effects are neglected a priori when configuring the model or the equipment for the testing. This is why the model must be carefully configured and the process tolerances known.

The agreement between simulation and testing improves with the quality of the measurement equipment, and also when the process has low tolerances. These solutions mean a considerable increase in the cost of the fabrication and testing, and therefore an increase of the final cost of the device. The designer must be aware that the results of simulations and experiments may not match exactly, but the agreement level between them must be very high.

An acceleration sensor is defined using the design values, keeping in mind the presence of tolerances. In Fig. 10.5, the structure of the sensor built by the simulator is shown. As can be seen, this accelerometer is composed of three parts: a comb-proof mass with four clamped suspensions, and two fixed electrodes.



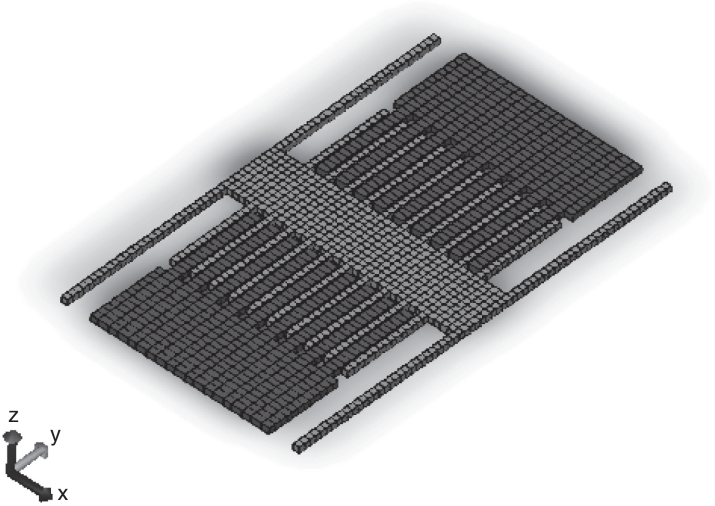
10.5 Solid model of the accelerometer sensor.

10.5.3 Mesh model

The solid model must be meshed in order to divide it into elements. The simulator uses the finite element method to perform the numerical simulations. It is recommended that the aspect ratio in the elements of the mesh should not be very high, because the simulation could provide inaccurate solutions or even fail if the problem to solve is not well-conditioned. Aspect ratios higher than 1:100 can be considered very high; aspect ratios between 1:1 and 1:10 are advisable.

The dimensions of the elements must take into account the dimensions of the different device regions. In the sensor in the example, there are elements with very different dimensions: the suspensions and the mass proof. Although the whole device can be meshed using unique element dimensions, the proof mass must be meshed using elements larger than those of the suspensions. This strategy decreases the number of elements in order to reduce the time simulation, since the more elements there are, the longer the simulation time will be.

The convergence analysis of the simulation is an important issue to take into account in order to ensure reliable results within a reasonable simulation time. This analysis involves reducing the dimensions of the elements and comparing the final results with those obtained with larger elements. If there is little difference between the simulations, the simulation with larger elements must be chosen in order to reduce the time simulation whilst retaining accurate results. The meshed solid model can be seen in Fig. 10.6.



10.6 Mesh model of the accelerometer sensor.

Table 10.4 Examples of BCs and InCs

Surfaces BC	Surfaces InC
Heat flux	Pressure
Voltage	Velocity
Load	Temperature
Volume BC	Volumen InC
Temperature	Temperature
Acceleration	Pressure
Heat generation	Fluid location

10.5.4 Boundary and initial conditions

Most of the problems are composed of behavioral equations, and boundary and initial conditions (BC and InC, respectively). In general, the simulator holds the majority of the equations and, in some cases, the simulator allows the introduction of new ones. However, the BC and InC must be always introduced by the designer.

These conditions could be applied to a surface or a volume of the meshed solid model, or at surface intersections or a point. Therefore, these surfaces, volumes and so on must be defined by names in order to set the conditions. In Table 10.4, some of the BC and InC are presented.

In this example, the sensor has only BC. The superficial boundary conditions are applied to the clamped side of the suspensions. In this case, there are no displacements of the faces, so this condition is imposed. Acceleration is applied to the whole device structure because gravity affects the beam and proof mass, and this condition is a volume BC. Finally, there are no initial conditions in the simulation because a steady-state analysis will be performed.

10.5.5 Results of the simulations

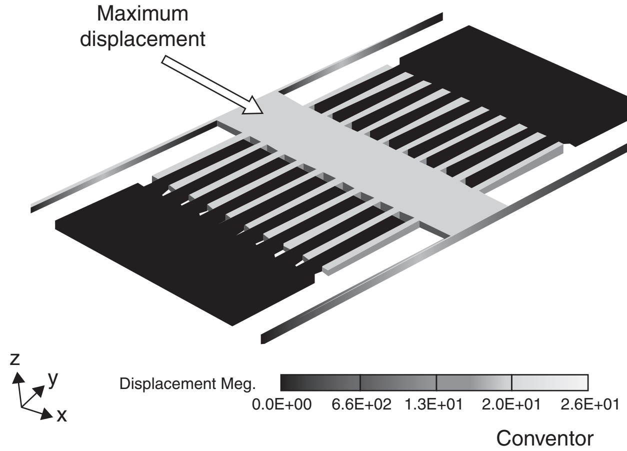
Once the solid model has been meshed and the BC and InC imposed, the simulations can be run. The simulator can provide the results as a table or a solid model mapped with colors as a function of a certain parameter values.

The results of the simulation depend on the simulation module used. Among others, the simulator uses mechanical (MEMMECH), electrical (MEMELECTRO) and piezoresistive (PIEZORESISTIVE) modules. Several modules can be used when performing microfluidics simulations; for example, BUBBLEDROPSIM, MEMSCFD and REACTSIM. The simulator allows the interaction between fluids and structures when using the module fluid structure interaction (FSI). In general, these software packages allow for simultaneous multiphysics simulations.

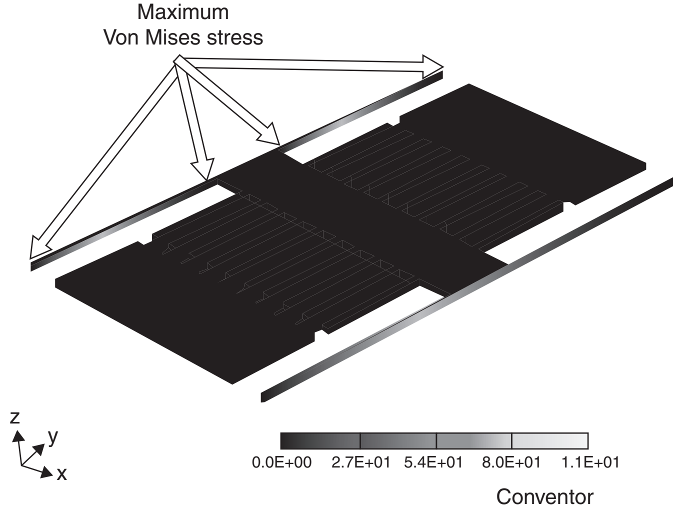
In the accelerometer example, the module MEMMECH is used in order to obtain the displacement as a function of the acceleration in the x-axis, and also to check the Von Mises stress to ensure that the structure does not fail during the operation. Then, the deformed structure is used to obtain the device capacitance with the MEMELECTRO module. The mechanical solution is shown for displacements in the x-axis, Fig. 10.7 and the distribution of the Von Mises stresses in Fig. 10.8. As can be seen in Fig. 10.7, the proof mass has the maximum displacement whereas the clamped are fixed. This displacement reduces the distance between the electrodes, and therefore increases the capacity of the device. The maximum stresses are located in the clamped beams and the joint of those beams and mass proof, as can be seen in Fig. 10.8. Finally, the electrical simulation provides the value of the capacity for both deformed and non-deformed microstructures.

The increase of capacity as a function of the acceleration can be obtained, and therefore the behavior of the device can be studied before fabrication. Also, the Von Mises stresses are useful because they are related to structure mechanical failure.

Parametric studies can be performed as a function of device dimension or mechanical properties. In this case, the proof mass displacements – and therefore the increase of capacity – can be studied; for example, as a function of the suspensions length in order to study the behavior of the structure and to find, if possible, an optimal configuration.


Maximum
displacement
z
y
x
Displacement Meg.
0.0E+00
6.6E+02
1.3E+01
2.0E+01
2.6E+01
Conventor

10.7 Representation of the displacement of the accelerometer microstructure when an acceleration is applied.


Maximum
Von Mises stress
z
y
x
0.0E+00
2.7E+01
5.4E+01
8.0E+01
1.1E+01
Conventor

10.8 Representation of the Von Mises's stresses of the accelerometer microstructure when acceleration is applied.

Once the device is simulated, if the results of these simulations showed a safe maximum of Von Mises stresses and a reasonable increase of capacitance for the acceleration which fulfills the device requirements, the sensor can be fabricated.

10.6 Conclusion

In this chapter, a general description of the most relevant materials and processes for microsystems has been presented. It aims to provide the basic information needed to understand the manufacturing design flow of a micro device, including the standard phases in industry: design, modeling, materials and process. This review includes basic descriptions, remarking on the main differences between MEMS technologies and standard manufacturing processes. The content of this chapter should be regarded as a starting point for a subsequent in-depth study.

10.7 References

Barlian, A., Park, W.-T., Mallon, J.R., Rastegar, A.J. and Pruitt, B.L. (2009). Review: Semiconductor piezoresistance for microsystems. *Proceedings of the IEEE*, **97**(3), 513–555.

Barth, P.W. (1990). Silicon fusion bonding for fabrication of sensors, actuators and microstructures. *Sensors and Actuators A*, **23**, 919–926.

Buddhadev, P.C., Ceyssens, F., De Moor, P., Van Hoof, C. and Puers, R. (2010). A high aspect ratio SU-8 fabrication technique for hollow microneedles for transdermal drug delivery and blood extraction. *Journal of Micromechanics and Microengineering*, **20**, 064006 (6pp).

Candler, R.N., Park, W.-T., Li, H., Yama, G., Partridge, A., Lutz, M. and Kenny, T.W. (2003). Single wafer encapsulation of MEMS devices. *Transactions on Advance Packaging*, **26**, 227–232.

Dargys, A. and Kundrotas J. (1994). *Handbook on Physical Properties of Ge, Si, GaAs and InP.* Science and Encyclopedia Publishers, Vilnius.

Deal, B.E. and Grove A.S. (1965). General relationship for the thermal oxidation of silicon. *Journal of Applied Physics*, **36**(12), 3770–3778.

Gogolides, E.,Vlachopoulou, M., Tsougeni, K., Vourdas, N. and Tserepi, A. (2010). Micro and nano structuring and texturing of polymers using plasma processes: potential manufacturing applications. *International Journal of Nanomanufacturing*. **6**(1), 152–163.

Hadjichristidis, N., Pispas, S. and Floudas, G. (2003). *Block Copolymers: Synthetic Strategies, Physical Properties, and Applications*. Hoboken, USA: John Wiley & Sons, Inc.

Hanneborg, A., Nese, M. and Ohlckers, P. (1991). Silicon-to-silicon anodic bonding with a borosilicate glass layer. *Journal of Micromechanics and Microengineering*, **1**, 139–144.

Heckele, M. and Schomburg, W.K. (2004). Review on micro molding of thermoplastic polymers. *Journal of Micromechanics and Microengineering,* **14** R1–R14.

Henmi, H., Shoji, S., Shoji, Y., Yoshimi, K. and Esashi, M. (1994). Vacuum packaging for microsensors by glass-silicon anodic bonding. *Sensor and Actuators A,* **43**, 243–248.

Kutchoukov, V.G., Laugere, F., van der Vlist, W., Pakula, L., Garini, Y. and Bossche, A. (2004). Fabrication of nanofluidic devices using glass-to-glass anodic bonding. *Sensors and Actuators A*, **114**, 521–527.

Laermer, F. and Schilp, A. (2006). Methods for anisotropic etching of silicon, US Patent #5, 501, 893.

Lee, T.M.H., Lee, D.H.Y., Liaw, C.Y.N., Lao, A.I.K. and Hsing, I.-M. (2000). Detailed characterization of anodic bonding process between glass and thin-film coated silicon substrates. *Sensors and Actuators A*, **86**, 103–107.

Lima de Faria, J. (1990). *Historical Atlas of Crystallography*. International Union of Crystallography by Kluwer Academic Publishers, Dordrecht.

Madou, M.J. (2002). *Fundamentals of Microfabrication and Nanotechnology*, 2nd edn, Boca Raton, Fl: CRC Press.

Maluf, N. (2002). *An Introduction to Microelectromechanical Systems Engineering*. Norwood, MA: Artech House.

Mata, A., Fleischman, A.J. and Roy, S. (2005). Characterization of polydimethylsiloxane (PDMS) properties for biomedical micro/nanosystems. *Biomedical Microdevices*, **7**(4), 281–293.

Miki, N., Zhang, X., Khanna, R. Ayón, A.A. Ward, D. and Spearing S.M. (2003). Multi-stack silicon-direct wafer bonding for 3D MEMS manufacturing. *Sensors and Actuators A*, **103**, 194–201.

Plößl, A. and Kräuter, G. (1999). Wafer direct bonding: tailoring adhesion between brittle materials. *Materials Science and Engineering: R: Reports*, **25**, 1–88.

Quinn, D.J., Spearing, S.M., Ashby, M.F. and Fleck, N.A. (2006). A systematic approach to process selection in MEMS. *Journal of Microelectromechanical Systems*, **15**(5), 1039–1050.

Schmidt, M.A. (1998). Wafer-to-wafer bonding for microstructure formation. *Proceedings of the IEEE*, **86**, 1575–1585.

Shimbo, M., Furukawa, K., Fukuda, K. and Tanzawa, K. (1986). Silicon to silicon direct bonding method. *Journal of Applied Physics,* **60**, 2987–2989.

Tadigadapa, S. and Mateti, K., (2009). Piezoelectric MEMS sensors: state-of-the-art and perspectives. *Measurement Science and Technology*, **20**(9), 30.

Tekin, T. (2011). Review of packaging of optoelectronic, photonic, and MEMS components. *IEEE Journal of Selected Topics in Quantum Electronics.* **17**, 704–719.

Wolffenbuttel, R.F. (1994). Low-temperature silicon wafer-to-wafer bonding using gold at eutectic temperature. *Sensors and Actuators A*, **43**, 223–229.

Wolffenbuttel, R.F. (1997). Low-temperature intermediate Au-Si wafer bonding; eutectic or silicide bond. *Sensors and Actuators A*, **62**, 680–686.

Xiao, Z.-X., Wu, G.-Y., Li, Z.-H., Zhang, G.-B., Hao, Y.-L. and Wang, Y.-Y. (1999). Silicon–glass wafer bonding with silicon hydrophilic fusion bonding technology. *Sensors and Actuators A*, **72**, 46–48.

11

Microactuators: design and technology

L. LI and Z. J. CHEW, Swansea University, UK

DOI: 10.1533/9780857099297.2.305

Abstract: Microactuators are emerging as important tools that are capable of performing the tasks of conventional tools in the macroworld, but they are much smaller and allow greater precision. A review of electrostatic, piezoelectric and electrothermal microactuators is presented. The basics of these three actuation principles and characteristics are briefly explained to highlight the motivation behind the selection of different microactuators. Various design models of electrostatic, piezoelectric and electrothermal microactuators are described and compared from the perspective of the mechanism and output performance. Limitations of existing designs are identified to improve some of these microactuators to gain better performance or reliability. Finally, key challenges in different microfabrication techniques associated with their process and compatibility with commercial foundries are highlighted, as this aspect decides the viability of microactuators commercially and economically.

Key words: electrostatic, piezoelectric and electrothermal microactuators, microelectromechanical systems (MEMS).

11.1 Introduction

Since the introduction of the integrated circuit (IC)-compatible surface-micromachining process, which enabled miniaturized systems that are composed of mechanical and electrical devices to be fabricated on a piece of silicon (giving birth to the term 'microelectromechanical systems' (MEMS) in the mid-1980s), remarkable progress has been witnessed in this new multidisciplinary field. Since the early 1990s, intensive research has been carried out in the MEMS field since they are a fundamental element in many micro and nano systems, and have a wide range of applications in many different disciplines, such as the optical field, the radio frequency field and the biological field. Knowing that the ultimate goal of MEMS research is to devise useful microstructures as a complete system, fundamental technological issues include materials, machining processes, micromechanical devices, and system integration have to be recognized. The advantages of MEMS lies in their miniaturized size (characteristics length ranging from 1 μm to 1 mm)[1] and low cost, which lends them to potential applications (e.g. in disposable, implantable, *in-vivo* medical equipment).

Table 11.1 Common microactuation mechanisms[7]

Input	Output (mechanical)
Electrical	Electrostatics Piezoelectricity
Magnetic	Magnetostatics Magnetostriction
Mechanical	Pneumatics Hydraulics
Thermal	Expansion Shape memory effect
Chemical	Phase change Combustion

Some technologies have already matured and are well-established within commercial markets, such as the micropump in ink-jet printing heads,[2] accelerometers in automotives for airbag deployment or in consumer electronics (including smart phones and games consoles) for controlling direction in their applications, free-fall sensors in hard disk drives of computers as well as gyroscopes in camcorders for blur image compensation.[3,4]

With MEMS technologies burgeoning at such a rapid pace, it is desirable to understand the main operational feature in order to further develop and improve the current designs, thus extending their use into further practical applications. Microactuators are regarded as the key element leading to the realization and physical functionalities of MEMS devices. Microactuators can accomplish complex tasks by arranging multiple subsystems in appropriate parallel and series combinations so that they work cooperatively, give higher output and perform tasks dexterously.[5] Microactuators are defined as actuators (with a scale of microns) produced using standard MEMS-based fabrication techniques including surface and bulk micromachining and other replication techniques.[6] These actuators convert (input) energy from one form – such as electrical, mechanical, thermal, magnetic, chemical and radiation energy – into the mechanical form (output). Common transduction mechanisms are listed in Table 11.1.[7]

Although there are many types of transductions, the basic idea behind making these microstructures work is quite straight forward: to generate force, which leads to simple mechanical work or movement such as bending, tilting and vibrating. Such force generation can be divided into external forces that are produced in the gap between fixed or stationary and movable parts by means of electrostatic[8] or magnetic energy,[9] and inner forces generated by materials with intrinsic actuation capabilities such as piezoelectric[10] and electrothermal materials.[11] Often, different mechanisms can be used to achieve similar results. For example, researchers have reported fine positioning of a read/write head in a hard disk drive using piezoelectric[12,13] electrostatic[14,15] and magnetic[16,17] methods, as well as moving optical structures such as a membrane deformable mirror, micromirrors or optical pinholes by means of electromagnetic,[18–22] electrothermal,[23,24] piezoelectric[25] and electrostatic[26–28] mechanisms.

The focus here is on microactuators that have electrical energy such as voltage or current, or both sources, as their input. This is because the power source is compatible with the power supply of electronic devices, which benefits from the well-developed field of electronic and electrical engineering, thus making the integration of mechanical systems and electronic circuitry on a single piece of chip simpler and more viable. Actuators that usually have an electrical signal as an input include electrothermal, piezoelectric, electrostatic and magnetic mechanisms. More details on these mechanisms will be discussed in Section 11.2.

11.2 Driving principles for actuators

As mentioned earlier, different methods of actuation can achieve the same results. However, there are several considerations that must be taken into account when making the choice of which actuation principle is to be implemented. Each actuation principle has its own advantages and disadvantages. The choice and the optimization of an approach should be made according to the requirements of a particular application. Attributes of electrostatic, piezoelectric, electrothermal and electromagnetic actuators will be examined to determine selection criteria for which mechanism to use.

11.2.1 Considerations in mechanisms selection

While the same material constant or physical law applies in the micro and macro world, the effect of these parameters on the whole system and its performance changes dramatically when scaling macroactuators down to microactuators. When the scale changes by 1/100, the volume is reduced by a factor of $(1/100)^3$, which is a million times less than before; meanwhile, the gravitational force changes by $(1/100)^4$, making the gravitational pull a trillion times less.[29] With negligible gravitational force, van der Waals, electrostatic and capillary forces come into play,[30] bringing out new parameters that are not available in the macroscopic world. For example, van der Waals force is sufficient to secure objects when using a micro-manipulator.[31] The selection of actuation principles is influenced by the enabling technology, structural dimensions, dynamic response, force or torque, displacement and power consumption.[32]

Size and physical properties

Small actuator dimensions enable tools to interact with components at a microscale or even nanoscale. However, sensible dimensions are essential for practical reasons. For example, the gap between the arms in a parallel beam is 1–2 μm, in order to achieve the effective manipulation of 1D nanomaterials (e.g. the alignment of carbon nanotubes on a surface).[33] Besides, designers have to comply with foundries' design rules in their plans.[34,35] Reduction in size of a mechanical system generally increases the structural stiffness, even at a thickness of one micrometre,[36] and its shock resistance.[2,37] Heat transfer mechanics such as transient effects and high spatial gradient also exhibit significant differences after having been scaled down.[38,39] This is especially crucial in electrothermal microactuators, where slight changes to structure dimensions will alter performance.[11,40] Besides mechanical properties, there are consequences for electrical properties: the break down voltage in very close gaps increases according to Paschen's Law.[41]

Output force and displacement range

The type of application affects the range of motion and applied force. Handling objects or specimens that are fragile (e.g. bio- and nanomaterials) requires a very low force, ranging from nanonewtons to micronewtons. Kim *et al.*[42] demonstrated force-controlled single cell manipulation using a microgripper in aqueous media with nanonewton force. On the other hand, applications such as high voltage large air gap switches require large deflection and forces up to several hundred microns.[43]

Actuation resolution and sensing

In the microworld, every single micrometre or nanometre is important. Nanoscale measurements are required to test microstructures such as nanowires.[44] High precision electrostatic impact microactuators that are able to generate stepwise motion of 10–20 nm[45] and piezoelectric microactuators with displacement magnitudes of 12.3 nm[10] have been reported. Such ultra-precise positioning is indispensable in applications such as X–Y stages, optical devices or the more demanding hard disk drive track positioning.[46] Measuring equipment needs to be able to detect minute changes in order to keep track of the microactuator's position in a microscale environment. Microactuators with a sensitivity of around 13.44–33.3 pC/μm were presented by Dong *et al.*[47]

Fabrication and material selection

Fabrication technologies are a crucial determinant of material selection, with silicon still playing the leading role. Ease of fabrication and compatibility to standard ICs would enable microactuators to be mass-produced at a low cost. There is no one material that is perfect for all types of actuation. For example, fast response materials might require high actuation voltage, whilst slow response materials need low actuation voltage. In high-speed applications[48] or switches,[43,49] if a fast response is desired, materials such as diamond, silicon carbide, alumina, silicon nitride and silicon are good candidates.[50] If a large force is required, titanium carbide, tantalum carbide or tungsten carbide can be used. However, using non-IC-standard materials will introduce additional costs and fabrication complexity. Hence, a trade-off is inevitable when designing a microactutor.

Power consumption

Power efficiency is a major concern in the macroscopic world, especially for portable devices. There is no difference when it comes to MEMS applications.[51,52] Low-power or energy efficient microactuators that operate at lower working temperatures will reduce nonlinear effects such as plastic deformation,[53] and hence increase the lifespan.

11.2.2 Electrostatic systems

Electrostatic forces are most commonly observed in the macroscopic world in daily activities such as combing hair. Such energy has been successfully harnessed in microactuators and it is the most frequently applied actuation due to its versatility,[54] simplicity and high compatibility with IC fabrication technology.[45,55] No special materials – for example, piezoelectric ceramics, or additional elements such as coils – are required, therefore eliminating some of the cost inefficient assembly process.

Electrostatic actuation

The basic working principle of an electrostatic system involves the concept and structure of the capacitor. It consists of charged, fixed and movable structures that are electrically isolated and move relative to each other. Energy, W, stored between the plates is given by:[7,32]

$$W = \frac{1}{2}CV^2 = \frac{1}{2}\frac{\varepsilon_r \varepsilon_0 abV^2}{d} \qquad [11.1]$$

where C is the capacitance, g is the gap or distance between plates, a, b are the lateral dimensions of the capacitor, V is the applied voltage and $\varepsilon_{r,0}$ are the dielectric constants of elements in the gap.

Derivations of W with respect to g and a yield the electrostatic forces, F:

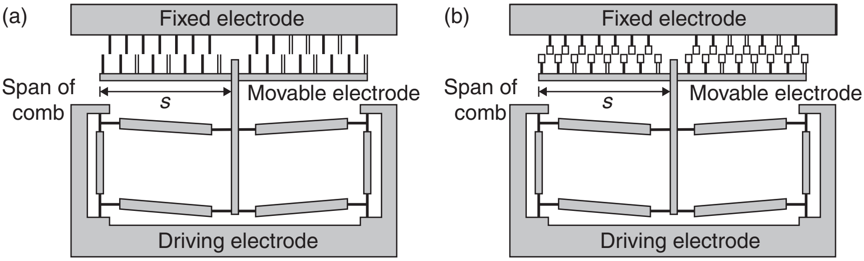
$$F_d = +\frac{1}{2}\frac{\varepsilon_r \varepsilon_0 \, abV^2}{g^2} \qquad [11.2]$$

$$F_a = -\frac{1}{2}\frac{\varepsilon_r \varepsilon_0 bV^2}{g} \qquad [11.3]$$

where F_d acts perpendicular to the capacitor plates and F_a acts in the direction of the capacitor plates. F_d is inversely proportional to the squared distance and therefore is only significant for very small gaps. F_a is independent of the actual position of the structures if the dimensions are large compared with the region of interaction. This result indicates that electrostatic principles are useful in miniaturized devices as the electrostatic force is significant. When the force is constant, the displacement is dependent on potential differences between the plates.

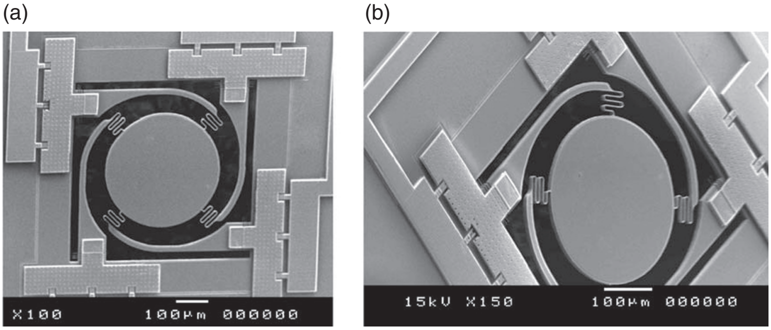
Common features

The most commonly adopted design is the comb-drive structure,[27,28,56–59] where two arrays of interdigitated (IDT) fingers – one fixed and the other freely movable – act together to produce an electrostatic force (see Fig. 11.1). Another type is known as 'gap closing'; it consists of two parallel-plate structures (see Fig. 11.2).[36,60] Electrostatic actuators are often characterized by low-power consumption[61] and a high switching speed.[48] Reduction in volume is easy for electrostatic microactuators because the forces available are only coupled to surfaces. Due to the small critical dimensions within the device – especially that of the gap between stationary and moving parts

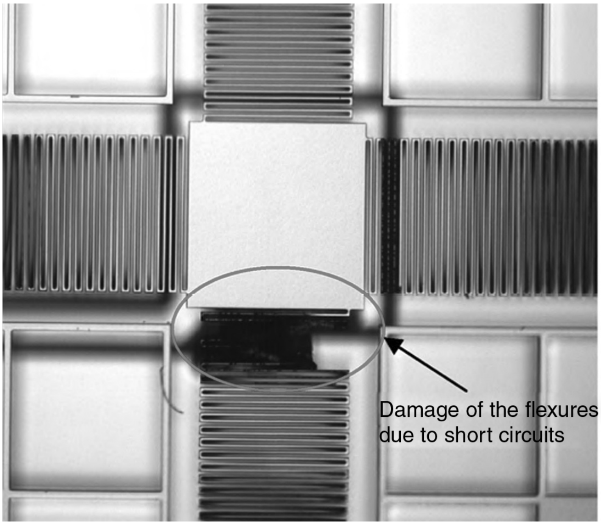


11.1 Electrostatic microactuator with: (a) uniform comb structure;[56] (b) spade-like comb structure.[59]

that is normally in the micron range – problems may arise from applications within real media. This is due to foreign particles such as dust impeding proper operation and thus degrading the performance of the device, or even causing failure if the debris creates a short circuit (see Fig. 11.3).[57] Actuation voltages reported so far range from 20 to 150 V.[57–59,61–63] Thus, only small deflections can be achieved at a low voltage, which is restrictive in some applications. Electrostatic actuators also suffer from a problem known as 'pull-in', where the movable parts stick to the fixed parts when actuation voltages reach the 'pull-in voltage'.[63] Due to their small footprint, which ranges from 0.1 to 1 mm^2, they are used in nanoscale applications,[56]



11.2 SEM images of electrostatic micromirror with gap-closing actuation (a) top view; (b) tilted view.[36]



11.3 Collapse of serpentine flexure due to short circuit between electrodes.[57]

performing tasks within a chip such as fine precision positioning[45] or *in situ* measurement.[64] However, *in situ* use in electron microscopes is discouraged as the electric field may interfere with the scanning electron beam due to the high input voltage. Fabrication of electrostatic microactuators with surface micromachining and using silicon-on-insulator (SOI) wafers[59] or polysilicon makes them compatible with the electronics interface.

11.2.3 Piezoelectric systems

Piezoelectric materials such as lead zirconate titanate (PZT) and polyvinylidene fluoride (PVDF) are usually employed as the transducers.

Piezoelectric actuation

Piezoelectric materials are physically deformed when an electric field is applied to them. Conversely, these materials produce electrical energy when are subjected to mechanical strain. These relationships can be described by:

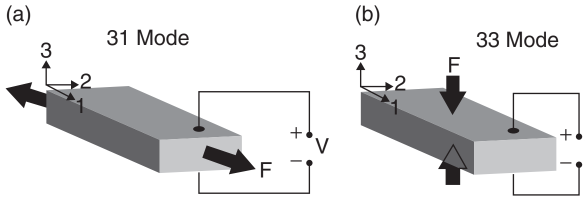
$$S_I = S_{IJ}{}^E T_J + d_{jI} E_J \tag{11.4}$$

$$D_i = d_{iJ} T_J + \varepsilon_{ij}{}^T E_j \tag{11.5}$$

where S is the mechanical strain, T is the mechanical stress, s is the compliance, d is the piezoelectric strain coefficient, E is the electric field, D is the electric charge density displacement, ε is the dielectric constant of piezoelectric material, I, J are 1, 2, 3, 4, 5, 6, i, j are 1, 2, 3 denoting the x, y, z directions.

Note that without the piezoelectric strain coefficient (d), Equations [11.4] and [11.5] are merely Hooke's Law and Gauss's Law, respectively. Thus, the performance is dependent upon the materials; the piezoelectric strain constant acts as a medium for the transduction mechanism to work. S, T, D and E are vectors. The presence of an electrical field across the material has an implication on its physical properties, while the mechanical stress has an effect on the permittivity of the material. For example, microactuators are excited by an electric field, usually a voltage source, and can cause the piezoelectric material to deform according to the direction of the applied voltage, thus inducing mechanical forces.

Common excitation modes are the 33 mode and 31 mode, as shown in Fig. 11.4. In the 33 mode, both the voltage and the force act in the 3 direction, while the 31 mode operates with the voltage acting in the 3 direction and the force being exerted in the 1 direction. The direction of the voltage can be manipulated by poling the material. If a voltage, V, is applied to a stress-free



11.4 Operating mode of piezoelectric beam: (a) 31; (b) 33.

piezoelectric material in the 3 direction, $T = 0$ and $E = \Delta l = \delta v\begin{bmatrix}0 & 0 & E_3\end{bmatrix}$, Equation [11.4] can be simplified as:

$$S_1 = d_{31}E_3 \qquad [11.6]$$

The displacement ΔX in the 1 direction can be expressed as:

$$\frac{\Delta X}{X} = S_1 = d_{31}E_3 \qquad [11.7]$$

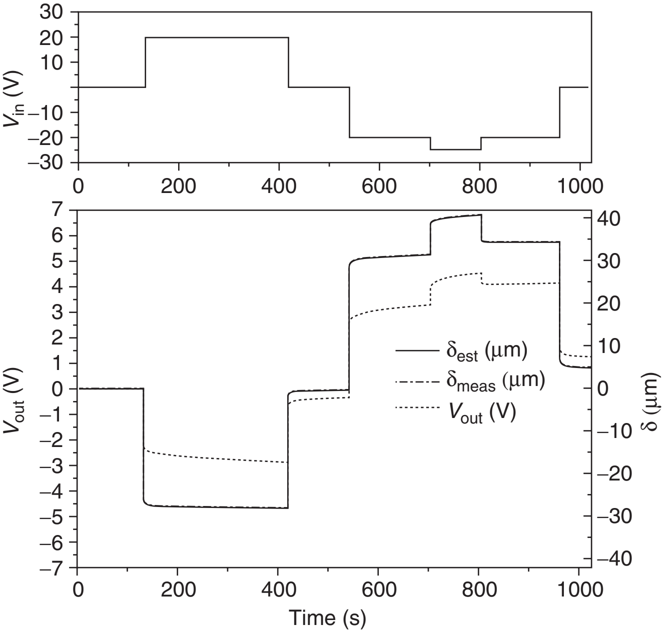
where X is the piezoelectric ceramic length in the x-direction. $E_3 = V/X$, hence

$$\Delta X = d_{31}V \qquad [11.8]$$

where variation of length in the y and z direction can be obtained in a similar fashion.

Common features

Like their electrostatic counterpart, piezoelectric microactuators are often deployed in high-speed applications due to their rapid response time (see Fig. 11.5)[65] and low-power requirement.[66] Large forces (i.e. hundreds of micronewtons) are demonstrated even in simple designs such as a thin film[66] or a beam[67] with an excitation voltage of 20 V. While conventional piezoelectric microactuators produce displacement of only 35 nm to 1.5 μm[31,68–70] with a high actuation voltage (from 50 V to several hundred volts) and have a large lateral dimension[10,71] or footprint (usually measured in mm^2) (see Fig. 11.27), a considerable number of improvements have been reported in recent years. Those findings showed that piezoelectric microactuators are quite versatile: small step piezoelectric microactuators can be employed in fine-positioning



11.5 Response of piezoelectric microactuator where the upper image shows the applied voltage and the lower image shows the output (displacement and voltage). The microactuator reacts almost instantly.[65]

applications, while large displacement microactuators can be integrated into devices such as micromirrors. Fabrication of piezoelectric ceramics is usually carried out at temperatures between 350°C and 700°C,[66,67] with some going over 1000°C[72,73] for several hours. Such temperatures are too high[74] for the integration of piezoelectric microactuators with post-complementary metal–oxide–semiconductor (CMOS) fabrication, since the component will be damaged at the elevated temperature. Prefabrication of PZT is difficult as it poses a contamination risk to the tools, which are also used in CMOS fabrication processes,[75,76] and silicon substrates can be contaminated as a result of lead diffusion.[77] Carrying out separate fabrication processes is inefficient in terms of time and cost. New materials, such as aluminium nitride (AlN),[78] are being investigated to decide if they can be fabricated using the conventional micromachining method.

11.2.4 Electrothermal systems

Thermal management is important to everything in the world, whether it is living or inanimate. IC is no exception; from the transistor level[80,81] to the complete system level[82] after chip level,[83] and from the package level[84] to the printed circuit board (PCB) level,[85] thermal effects change the electrical properties and this can affect the circuitry. However, with proper design and control, extreme phenomena observed in ICs (such as buckling and expansion as a result of the different thermal expansion coefficients[86] of silicon and metal) have shown the capability of miniaturized systems to generate large forces, which has been incorporated as one of the electrothermal microactuators' mechanisms. Electrothermal actuation is therefore preferable to piezoelectric and electrostatic mechanisms when large forces and deflections are required.

Electrothermal actuation

Electrothermal actuators operate on the principle of Joule heating and differential thermal expansion.[87] An electrical current flows through the actuator, which is usually an electrical closed loop, and heat is generated. The heat generated in an ohmic conductor is expressed by:

$$Q = j^2\rho = I^2R \qquad [11.9]$$

where j is the current density vector, ρ is the specific electric resistivity, I is the current passing through the materials, R is the materials' resistance and Q is the generated heat (power) per unit volume. A general thermal expansion is given by:[88]

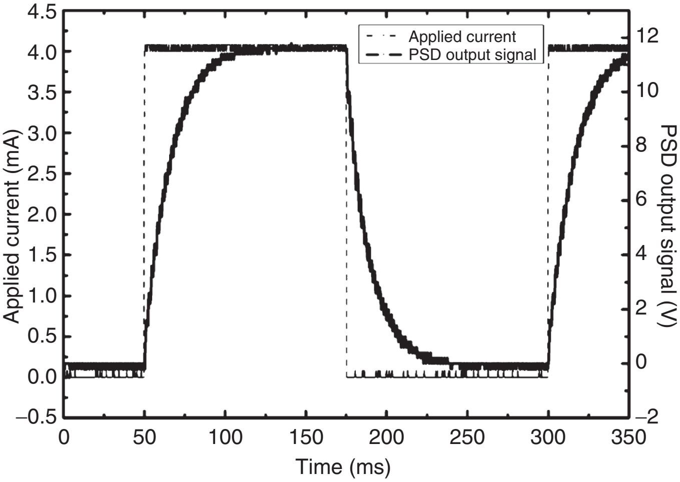
$$l = l_0\left(1 + \alpha\Delta T\right) \qquad [11.10]$$

where l is the final length, l_0 is the initial length, α is the thermal expansion coefficient and ΔT is the temperature change.

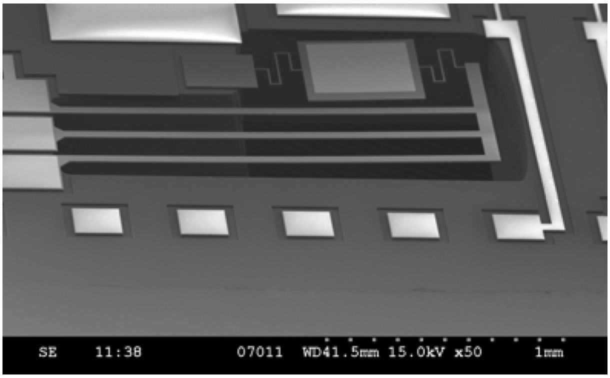
From Equation [11.10], it can be seen that Joule heating in electrothermal actuators is heavily dependent on the properties of materials (e.g. the physical properties, the electrical properties, the thermal properties and the volumetry of the actuators). This indicates that careful selection of materials and geometry is crucial for the optimal performance of electrothermal microactuators. Joule heating causes temperature changes within the actuators, and the temperature variation leads to differential thermal expansion.

Common features

Electrothermal microactuators have a fairly slow response time compared with piezoelectric and electrostatic microactuators: the transducers require time to heat up and cool down, thus limiting their operating frequency to less than 1 kHz. Operation at frequencies beyond the response time will confine the actuator to one position.[89] Li *et al.*[11] reported that the device reacts 17.3 ms after the voltage is applied, which is about 1000 times slower than piezoelectric actuators. Increasing input power reduces the bandwidth because the actuators heat up to a higher temperature and thus take more time to return to their initial temperature. Electrothermal actuators also use large amounts of power, with the power requirements ranging from several milliwatts[11,89–92] to watts,[93] at several volts and several milliamps of current (see Fig. 11.6). However, electrothermal actuators are capable of exerting



11.6 Response time of an electrothermal microactuator with large vertical displacement.[91]



11.7 SEM image of an electrothermal MEMS scanner actuated by a single structural layer, three-beam electrothermal out-of-plane actuator.[99]

high output forces[94] and they produce large displacements at a fairly low excitation voltage (< 10 V), which compensates for their weakness. For example, with only 5.3 V, vertical displacement of 0.62 mm[91] and more than 3 mN[95] can be achieved. Many papers have reported the use of the services of Poly Multi-User-MEMS-Process (MUMPs) that fabricates their devices.[11,89,92,96] PolyMUMPs is a commercial foundry and the industry's longest-running MEMS multi-project wafer service, working with systems that are compatible with current IC manufacturing processes.[97] Design configurations presented so far have included U-shaped (see Fig. 11.7),[35,40] multimorph (more commonly known as 'bimorph'),[91] and V-shaped (chevron).[98]

11.3 Design and analysis of microactuators

Various common microactuator designs are presented in this section, including comb-drive, parallel-plate, scratch drive actuators (SDAs), U-beam, V-beam, multimorph, compliant mechanisms, embedded skeleton-based polymeric actuators, membrane or thin film, plus some other configurations. Challenges associated with modeling and analyses are examined.

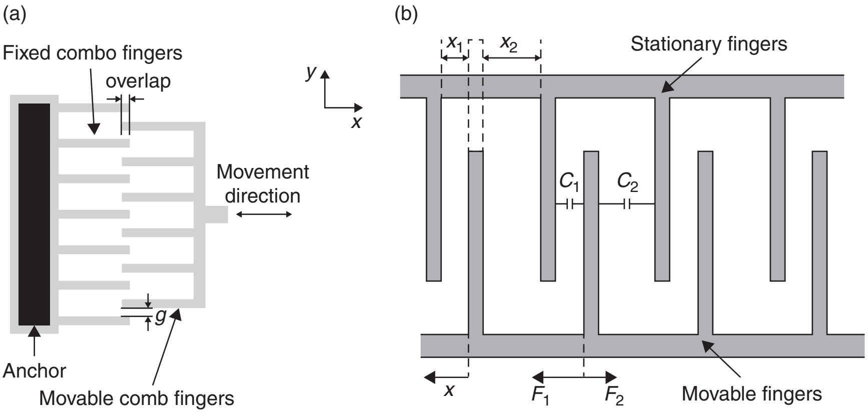
Also, the basic operation of microactuators will be discussed, together with comparisons and optimizations of the microstructures in these mechanisms.

11.3.1 Comb-drive actuators

The motion of comb-drive actuators and parallel actuators is controlled by the equilibrium of the electrostatic force and spring force of the suspension system. During lateral motion or parallel actuation, the attractive force between two comb electrodes is mainly due to the fringing fields, since the small fingers are thick compared with their width and length.[100] Since the distance between the comb fingers is constant, capacitance changes linearly with regard to the area of the plates that overlap during such movement. Meanwhile, transverse movement of the plates maintains the overlapping area and changes the size of the gap between the comb fingers (see Fig. 11.8).[58] Keeping the direction of motion (in both cases) along the x-axis, the electrostatic forces are:

$$F = \frac{n\varepsilon t_e}{g} V^2 \qquad [11.11]$$

$$F = \frac{1}{2} n\varepsilon A \left[\frac{(x_1 + x_2)(x_2 - x_1 + 2x)}{(x_1 - x)^2 (x_1 + x)^2} \right] V^2 \qquad [11.12]$$



11.8 Comb actuators: (a) lateral motion with force corresponding to Equation [11.11];[100] (b) transverse motion with acting force expressed by Equation [11.12].[58]

where n is the number of comb pairs, t_e is the thickness of the plate, g is the gap spacing, ε is the permittivity of the medium, A is the overlapping area of each finger pair, $x_{1,2}$ are the initial gap spacing, x is the changes in gap spacing and V is the voltage.

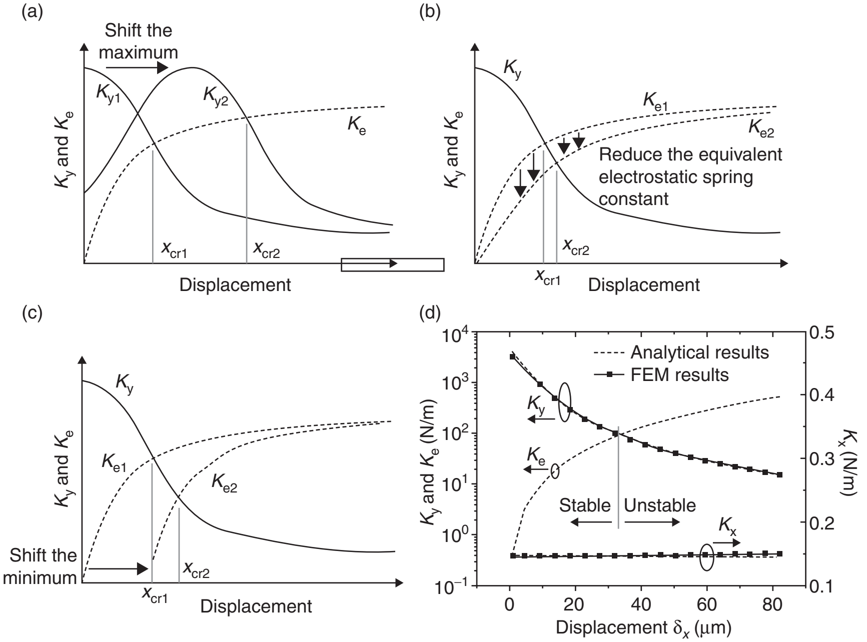
From both equations, increasing the number of comb pairs and increasing the voltage can increase the force output. Reducing the gap spacing generates a higher force for lateral motion actuation, while increasing the overlapping area gives more force to a comb-drive that is in transverse motion. The gap-closing actuator produced a larger force than parallel motion actuators.[58] Thus, large arrays of comb structures can be seen in many designs[27,57,58,101] since, as long as the size is within the design constraint limit, they increase the force. For a fixed area, thinner fingers with smaller gap spacing increase the power density since more comb pairs can be packed in. However, comb fingers that are too thin may deform individually and contact each other.[102] The minimum gap spacing between the fingers and the displacement are limited by pull-in or snap through instability.[103,104] This limitation is critical in parallel-plate actuators, as transversely driven actuators' displacement is restricted to about one-third of the gap spacing.[105,106] Once the driving voltage reaches the critical pull-in point, the microactuators will fail as their mobile comb structures stick to the fixed part.[107] Moving comb drives also experience vertical levitation due to unbalanced electrostatic fields.[108]

According to Hirano *et al.*,[104] a comb-drive actuator operates stably when $K_y > K_e$, where K_y is the spring stiffness in the y-direction and K_e is the negative spring constant. Generally, under normal conditions the maximum stable displacement $\delta_x^{\max}$ is expressed by:

$$\delta_x^{\max} = \frac{1}{2}\sqrt{L_0^2 + 2\frac{K_y}{K_x}g^2} - \frac{L_0}{2} \tag{11.13}$$

This is a function of the spring stiffness in the y-direction K_y, the spring stiffness in the x-direction K_x, the length of overlap L_0, and the gap spacing g between the fixed and movable electrodes. Different methods for increasing the stable displacement have been reported, including changing the spring constants of the system by altering the maximum setting of the suspension system (Fig. 11.9a), reducing the negative spring constant (Fig. 11.9b) and shifting the minimum setting of the negative spring constant (Fig. 11.9c).[109] The mechanical design of the suspension system is therefore crucial for large displacement comb-drive actuators. From Equation [11.11]–[11.13], it can be observed that the driving force does not change with motion, so for compliance in the x-direction, linear spring stiffness K_x is desirable, whilst a high stiffness ratio K_y/K_x is required to avoid pull-in instability and to increase displacement.

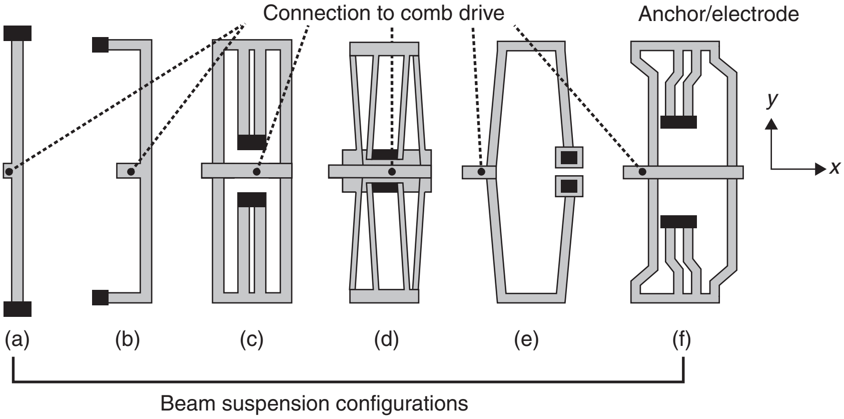
Proposed designs include the clamped-clamped beam, which provides a high stiffness ratio K_y/K_x, although stiffness K_x is linear only in the small



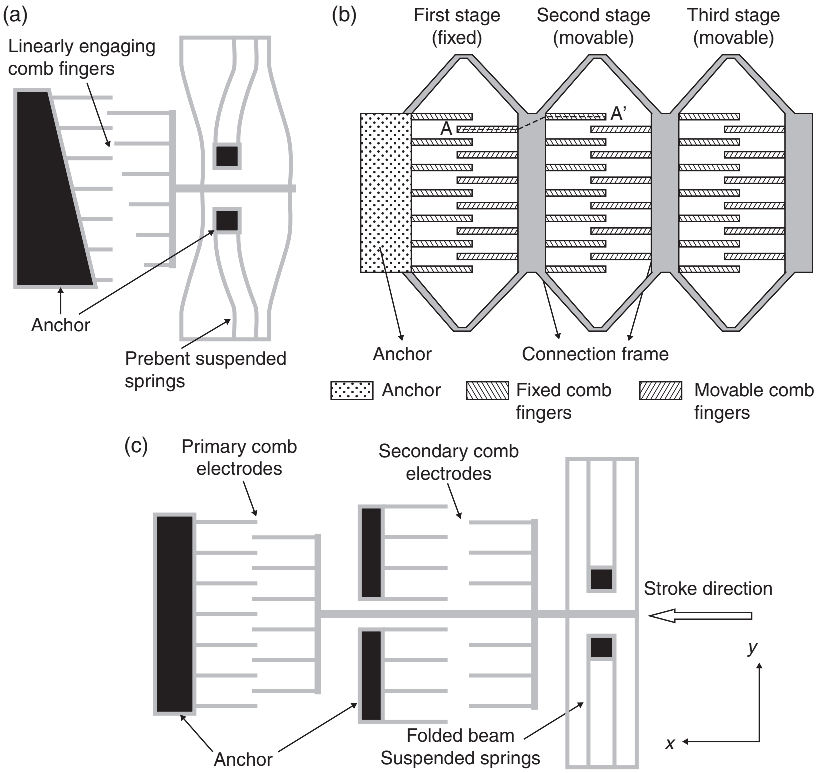
11.9 Increment of displacements: (a) maximum of spring constant shifting; (b) equivalent electrostatic force reduction; (c) minimum of spring constant shifting;[109] (d) stable and unstable operating region of a tilted folded beam.[103]

deflection range. Crab-leg flexure improves the linearity of spring stiffness K_x, but the stiffness ratio reduces at higher displacements.[110] The folded-beam design is more practical. However, K_y decreases at higher displacements.[54,110] In improved folded-beam designs, there were introductions of tilted folded beams,[103] prebent-tilted beams[111] and hybrid springs with folded beams (see Fig. 11.10),[112] which shifts the maximum of spring stiffness K_y and improves the displacement. Besides the parameters relating to thickness and number of comb pairs, linear variation of the comb finger length can enhance the side stability by reducing the overlapping area that affected the side electrostatic force.[111] Varying the gap spacing also improves the displacement and reduces driving voltage.[104] Additional secondary comb structures are capable of extending the displacement by reducing the minimum equivalent electrostatic spring constant, as shown by Hou *et al.*[109] Chiou *et al.*[113] have also demonstrated a 200% increase in displacement by using three cascading comb drives (see Fig. 11.11b). On the other hand, Borovic *et al.*[114] proposed the use of active feedback to counteract pull-in.

Displacement in parallel-plate comb actuators can be enhanced by geometry leverage,[115] the two-beam-method,[116] employing series capacitors for negative feedback,[117] current drive,[118–120] voltage drive,[121] feedback linearization[107] and finite time stabilization[122] to break the one-third gap limit and use 65–90% of the available gap. Levitation suppression is attained by alternating the electrodes on every comb finger with a striped ground plane that is positioned underneath the comb structure, and designing structures with vertically stiff suspensions. Levitation can be carried out in a controlled manner by using soft suspensions with differential and common mode voltages



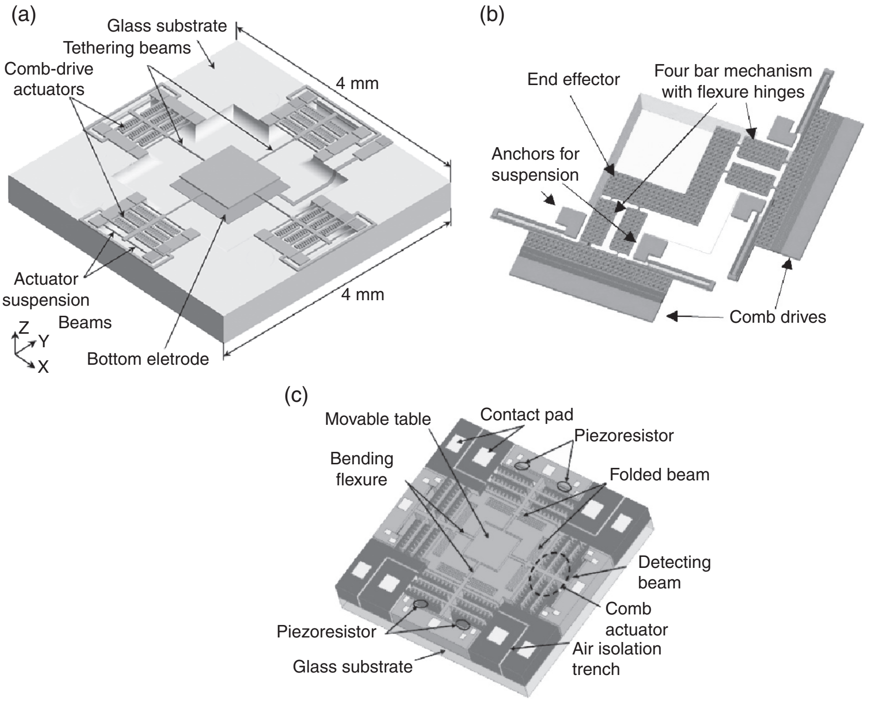
11.10 Laterally driven comb-drive with different suspension systems: (a) clamped-clamped beam; (b) crab-leg beam; (c) folded beam; (d) tilted folded beam; (e) hybrid spring; (f) prebent-tilted beam.


(a)
Linearly engaging comb fingers
Anchor
Prebent suspended springs
(b)
First stage (fixed)
Second stage (movable)
Third stage (movable)
A
A’
Anchor
Connection frame
Anchor
Fixed comb fingers
Movable comb fingers
(c)
Primary comb electrodes
Secondary comb electrodes
Stroke direction
y
x
Anchor
Folded beam Suspended springs

11.11 Displacement enhancement: (a) linearly engaging teeth with prebent suspension; (b) three-stage cascading comb drives;[113] (c) sequential engagement of primary and secondary comb drives by voltage switching.[109]

that are applied to the two electrodes.[108] However, some designs use this levitation force to move vertically.[101]

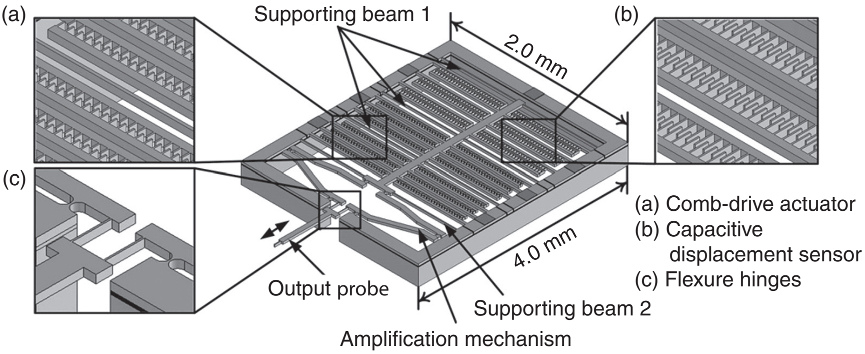
For applications such as the nanopositioning of micromirrors or in stages that require a higher degree-of-freedom (DOF), four actuators are usually positioned around a central stage, suspended by a folded-beam or a clamped-clamped beam that provides the restoring force. Direct connection of the stage to the comb-drive will limit the displacements and be more susceptible to side instability. Therefore, tethering beams with an appropriately low stiffness which connect the stage to the comb-drive allows orthogonal movement and decreases crosstalk between axes. Lower stiffness, however, leads to end-effector rotation, lower resonant frequency, lower bandwidth, and complex dynamics.[123] Parallel kinematics mechanisms (PKMs) with comb-drive actuators, as shown in Fig. 11.12b, provide an increased motion range, high structural stiffness and a balanced mechanical structure.[123] Besides


(a)
Glass substrate
Tethering beams
Comb-drive actuators
4 mm
Actuator suspension
Beams
4 mm
Bottom eletrode
Z
Y
X
(b)
End effector
Four bar mechanism with flexure hinges
Anchors for suspension
Comb drives
(c)
Movable table
Contact pad
Piezoresistor
Bending flexure
Folded beam
Detecting beam
Comb actuator
Piezoresistor
Air isolation trench
Glass substrate

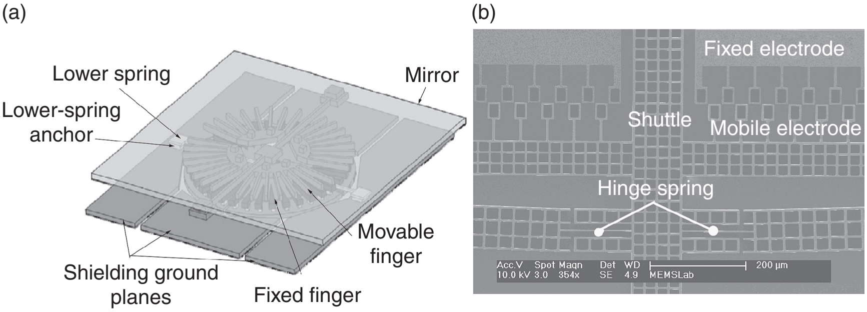
11.12 Nanopositioning stages with comb-drive actuators: (a) three-axis nanopositioning stage with z-axis motion induced by parallel-plate actuator;[126] (b) parallel kinematic four-bar mechanism with flexure hinge configuration;[123] (c) an XY stage with piezoresistor sensor.[125]

producing motions, IDT comb structures can also be used as sensing capacitors.[42,124] However, parasitic capacitance has the potential to overwhelm the signal. A piezoresistive feedback system can be integrated during the nanopositioning stage in order to resolve this problem.[125]

Nanoelectromechanical systems (NEMS) further push the envelope of microactuator miniaturization, as well as advocating the ease of integrating electrostatic microactuators with IC. Comb-drive actuated manipulators can convert the micrometre motion into subnanometre step displacement with high repeatability and force output, and with a linear amplification mechanism. A nanomanipulator with dimensions of 1 mm with a 0.15 nm resolution, ±2.55 μm motion range and 98 μN force capability has been presented by Liu *et al.*[124] (see Fig. 11.13). Other designs include combinations of linear and vertical comb-drive actuators with linear and torsional springs, which is challenging in terms of fabrication.[127] Kwon *et al.*[59] (see Fig. 11.14b) demonstrated comb actuators with spade-like fingers that operate in bi-stability with one input pulse voltage. The comb finger tips are made broader


(a)
Supporting beam 1
(b)
2.0 mm
(c)
(a) Comb-drive actuator
(b) Capacitive displacement sensor
(c) Flexure hinges
4.0 mm
Output probe
Supporting beam 2
Amplification mechanism

11.13 Solid model of nanomanipulator with an amplification mechanism driven by comb actuators.[124]

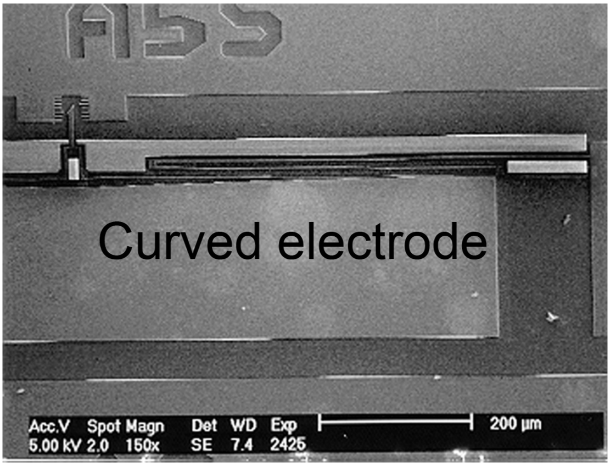

(a)
(b)
Lower spring
Lower-spring anchor
Mirror
Movable finger
Shielding ground planes
Fixed finger
Fixed electrode
Shuttle
Mobile electrode
Hinge spring
Acc.V Spot Magn Det WD 200 µm
10.0 kV 3.0 354x SE 4.9 MEMSLab

11.14 (a) A two-axis MEMS scanner with radial vertical comb-drive actuators;[127] (b) an electrostatic microactuator with spade-like comb structures for bi-stable latching.[59]

so that when movable combs are between the fixed combs, the smaller gaps between the tips means that the outward force is stronger. Also, Tee *et al.* proposed the use of adaptive controls for a 1 DOF electrostatic microactuator system so that it could be driven bidirectionally.[128]

11.3.2 Parallel-plate actuator

This configuration works in a similar way to the transverse mode comb actuators, except that it has only one pair of plates instead of a large array. It also suffers from pull-in instability and displacement of one-third of the gap spacing. The solutions to these problems are the same as the solutions to problems in parallel-plate comb actuators. Parallel-plate actuators are often employed to move a micromirror or three-axis stage vertically while horizontal movements are actuated by comb drives.[57,126] Curved electrode actuators have been proposed for use in situations where the gap is small

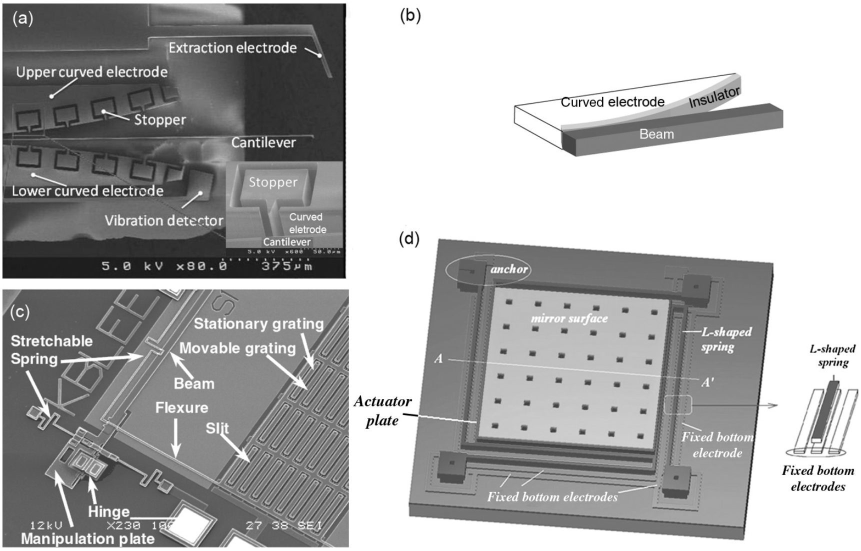


11.15 An optical microswitch with curved electrode structure.[26]

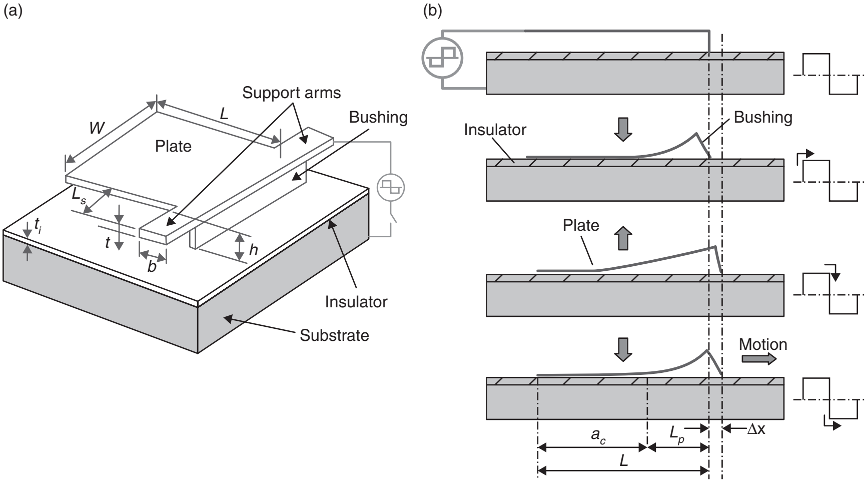
near the anchored edge of the electrode and gradually increases along the electrode to obtain larger displacement than conventional parallel-plate actuators (Fig. 11.15).[129,130] An insulator layer[129] or bumper (stopper)[130] is formed along the curved electrodes. A recent design is an L-shaped spring for electrostatic actuated deformable micromirrors (see Fig. 11.16d),[131] which increases the deflection that has been limited by the gap spacing due to the commercial PolyMUMPs process. Also, slit and grating actuators, which have an opening on the movable plate and fixed electrodes gratings below the slits, reduce pull-in stability, extend travel range and lower actuation voltage.[63]

11.3.3 Scratch drive actuator

SDAs[132,133] comprise an L-shaped plate, bushing and support arms that are connected to a spring box. By applying a bipolar pulse voltage, the plate is attracted towards the substrate during the positive cycle. However, the end of the plate, which is attached to the bushing, cannot be pulled down and is therefore tilted, sliding forwards due to the warping of the plate. When the voltage goes to the negative cycle, strain energy in the plate will pull the actuator forward as a one-increment step. This type of actuator is capable of generating high forces (>200 µN), large displacement (~100 µm), and nanometre positioning resolutions (~10 nm). Its speed can range from 70 to 250 µm/s, depending on the amplitude and frequency of the input signal. However, minimum driving voltages of 60 V may be required and can go as high as 290 V.[132] Also, the upward–downward and impact motion that is caused by warping may not be suitable for gripping and precision nanoassembly


(a)
Extraction electrode
Upper curved electrode
Stopper
Cantilever
Lower curved electrode
Vibration detector
Stopper
Curved eletrode
Cantilever
5.0 kV x80.0 375µm
(b)
Curved electrode
Insulator
Beam
(c)
Stretchable Spring
Stationary grating
Movable grating
Beam
Flexure
Slit
Hinge
Manipulation plate
12kV X230 100 27 38 SEI
(d)
anchor
mirror surface
L-shaped spring
A
A'
Actuator plate
Fixed bottom electrode
Fixed bottom electrodes
L-shaped spring
Fixed bottom electrodes

11.16 SEM images: (a) curved electrode with stopper;[130] (b) curved electrode with an insulator layer;[129] (c) slit actuator;[63] (d) an L-shaped spring suspended micromirror.[131]


(a)
Support arms
Bushing
W
L
Plate
L_s
t_i
t
b
h
Insulator
Substrate
(b)
Insulator
Bushing
Plate
Motion
a_c
L_p
Δx
L

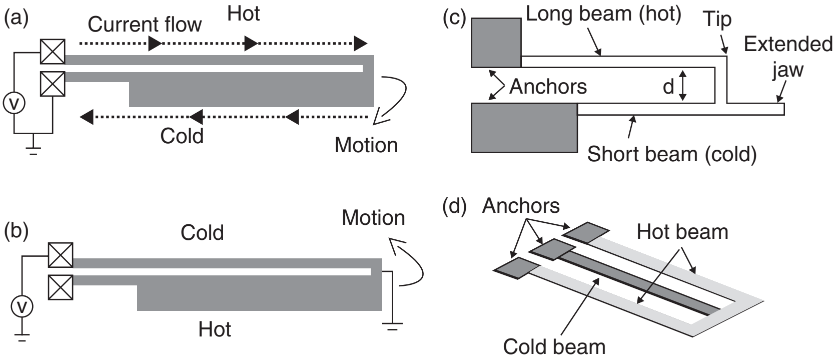
11.17 (a) 3D configuration of an SDA; (b) illustration showing warping motion of SDA.[133]

operations. SDAs are highly sensitive to plate length and applied voltage. A larger step size can be obtained by increasing the bushing height or the applied voltage. Driving voltages change linearly with the bushing height and exhibit an inverse relationship with the plate length (see Fig. 11.17).[133]

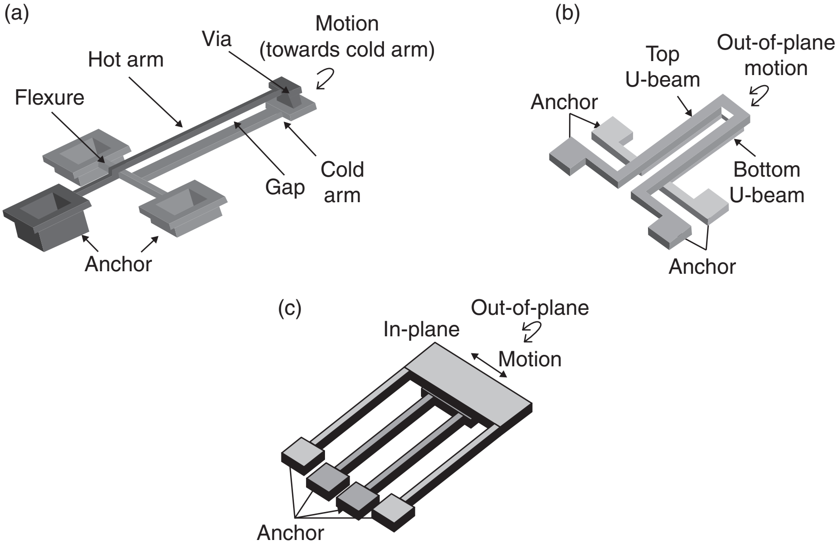
11.3.4 U-shaped (pseudo bimorph) beam

Also dubbed 'the Guckel',[134] this electrothermal actuator configuration is an electrically closed loop that resembles a 'U' shape. It generally consists of a 'hot' arm and a 'cold' arm that usually has a flexure in the in-plane U-beam. The hot arm has a higher temperature than the cold arm because it has a narrower conductive path, which increases the resistivity and current density. In-plane deflection towards the cold arm is achieved by the differential thermal expansion of the arms, giving it one operation mode ('open' or 'closed'). The flexure is a thin structure positioned between the cold arm and the anchor that amplifies the deflection. When a current is passed through the narrow and wide arms in parallel, as shown in Fig. 11.18b, the structure will deflect towards the narrower arm because the wider arm has a lower resistance, hence drawing more current and becoming hotter than the narrow arm. Another configuration is the long–short beam, which bends towards the shorter arm because the longer beam elongates more than the short beam when heated.

Out-of-plane unidirectional U-beams, as shown in Fig. 11.18d, are machined so that the initial stress of the material (e.g. silicon) causes the long, suspended cantilever to slightly bend upwards from the substrate. Subsequent heating causes expansion of the outer hot beam while being constrained by the centre cold beam, leading to further vertical movement. Unidirectional U-beams can also be made to have the hot arm positioned above the wider cold arm, which works on the same principle as in-plane U-beams.[137] Other designs have been reported to give operational directions. A bidirectional vertical actuator (BDVA), shown in Fig. 11.19b, is composed of two U-shaped arms (top and bottom) that are connected at the free edge.



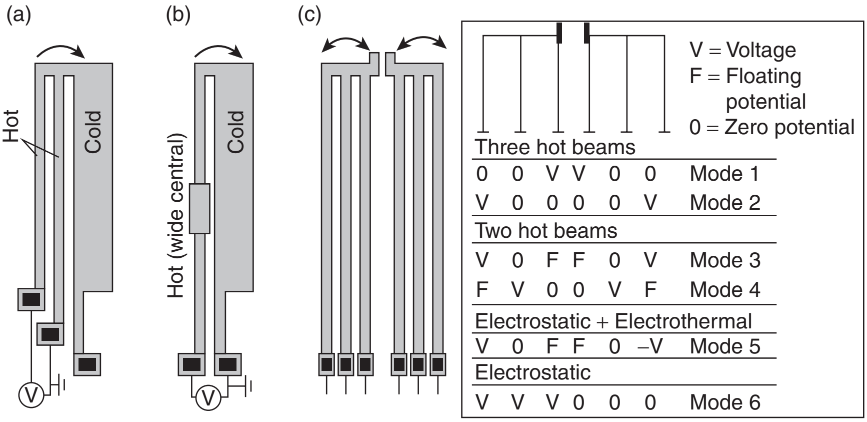
11.18 In-plane unidirectional U-shaped beams: (a) in series; (b) in parallel;[135] (c) long–short arms;[136] (d) out-of-plane unidirectional U-beam.[99]



11.19 Out-of-plane U-beams: (a) unidirectional; (b) bidirectional;[137] (c) with 2DOF.[35]

Passing a current through the bottom U-arm bends the tip upwards, while applying voltage across the anchors of the top U-beam deflects the actuator downward to the bottom arm. A 2DOF design that is able to deflect horizontally and vertically has been reported (see Fig. 11.19c).[35] It has four equidistant parallel arms that are anchored to the substrate at one end and connected to each other by a rigid central shuttle at the tip of the beam. The outer and inner arms are at different elevations, with the outer arms slightly higher. If a closed loop is formed appropriately with two of the arms, the deflection directions of the microactuator can be manipulated. For example, passing a current through the two beams on the right produces motion to the left, whilst exciting the outer arms gives rise to negative vertical motion.

As mentioned earlier, geometry is an influential factor in actuators' performance. The ratio of the cold arm length, L_c, to the hot arm length, L, should be around 0.85–0.9, which also means that a flexure of 10–15% of the total arm length and the wider cold arm are required for maximum deflection.[96,138] Using thicker arms for in-plane operating actuators increases the out-of-plane stiffness, thus avoiding some of the unwanted torsion of arms, the frequency dependent cross-coupling effects and misalignment in operations such as gripping nanoscale objects.[139] However, the width and the thickness must not be too great or the beam will be too stiff and cold to produce high deflection. Also, the beams must not be too long or there will be increased electrical resistance, and therefore more power will be consumed



Three hot beams						
0	0	V	V	0	0	Mode 1
V	0	0	0	0	V	Mode 2
Two hot beams						
V	0	F	F	0	V	Mode 3
F	V	0	0	V	F	Mode 4
Electrostatic + Electrothermal						
V	0	F	F	0	–V	Mode 5
Electrostatic						
V	V	V	0	0	0	Mode 6

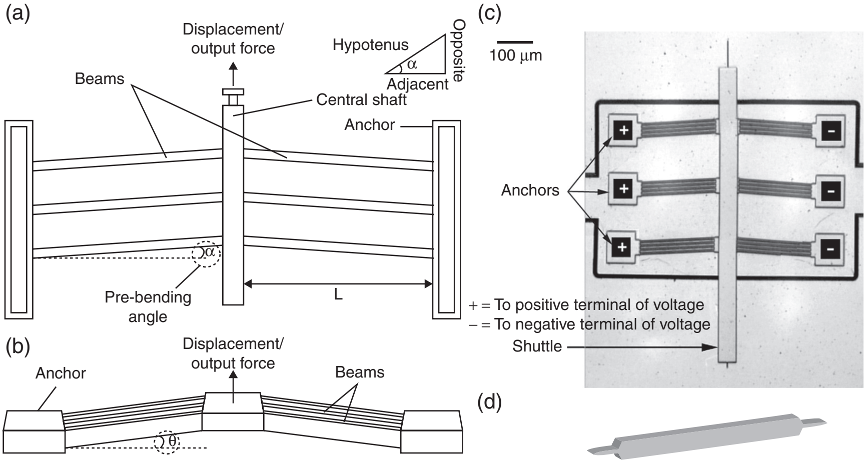
11.20 Top view of modified U-beams: (a) double 'hot' arm-U-beam;[40] (b) U-beam with widened hot arm at middle region of hot arm;[53] (c) a pair of U-beams with uniform arms that enable six different actuation modes.[33]

and less force will be produced due to bowing. Failure due to the stiction of the beams to the substrate during fabrication may also occur.[140] Also, peak temperatures measured at the middle of the hot beam[141] can be lowered by widening the central region of the beam (Fig. 11.20), improving the power-handling capability[53] and thus preventing failure due to localized melting or permanent plastic deformation. Although this modified structure leads to lower deflection than those with uniform arms at the same input power, it can operate at a higher power and produce higher maximum displacement than classical structures.

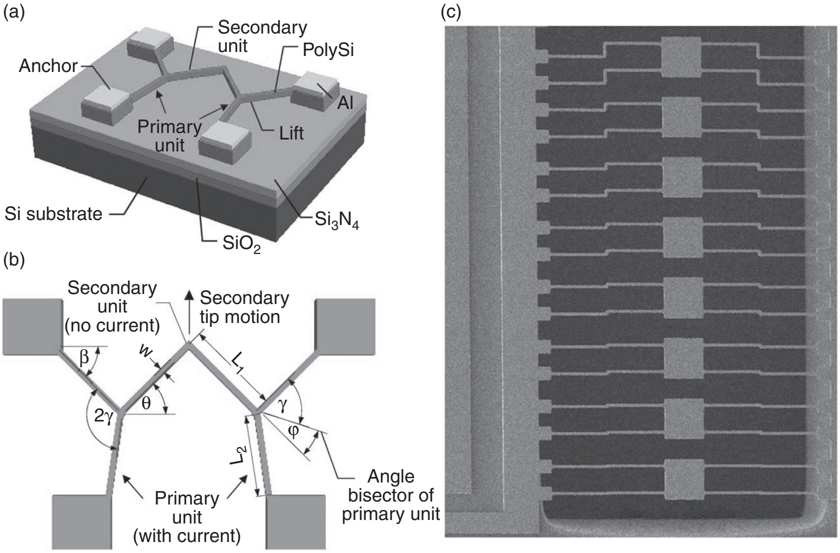
Large temperature differences between the hot and cold arms leads to a higher deflection.[142] Therefore, long–short beams with uniform arms perform poorly.[138] This can be rectified by adding another hot arm with a separated anchor in parallel with the existing hot arm and cold arm. Current only flows through the hot arms; the cold arm and flexure no longer conduct the current, resulting in a lower temperature and allowing the flexure to be narrower and thus more flexible. Consequently, a larger temperature difference and a higher deflection are attained. Also, deflection can be enhanced by depositing a highly conductive metal thin film, such as gold or aluminium, onto the cold arm.[96,138,143] This reduces the thermal resistance and the temperature on the cold arm to give a larger temperature difference, thus doubling the deflection. However, these thin films are damaged at very high temperatures.[143] Incorporation of diamond nanoparticles into nickel actuators has been shown to reduce power consumption by 73% and has improved performances by enhancing their properties.[144] Actuators built on trenched substrates can reduce power consumption since heat loss from the substrate is minimized.[89]

11.3.5 V-beam (chevron/bent-beam) actuator

The V-beam design features an array of bars that are attached symmetrically to either side of a movable central shuttle. The beams are designed with a prebend angle, α, that predefines the buckling direction. Joule heating causes the beams to expand and pushes the shuttle in the direction of the opposite side to angle, α. Within thermal buckling limits, longer beams can extend the travel range of the shuttle. More beams pairs produce more out force. Chevron actuators also provide flexibility by controlling the peak temperature and gripping force by increasing the beam's cross-sectional area. As with the U-beam, displacement of the V-beam actuator can be improved by varying the cross-sectional area by contouring the hot arms along the actuator's length, which provides more uniform temperature distribution along the length. Peak temperature and power consumption are therefore reduced. Also, contouring reduces thermal errors, thereby increasing the accuracy of the actuator that is required for nanopositioning applications.[53,145] Due to the processes that take place during fabrication, a parasitic angle, θ, is usually formed between the beams and the horizontal plane, as shown in Fig. 11.21b. Vertical motion of chevron actuators can be enhanced by increasing this angle.[146] Figure 11.22a and 11.22b shows another bent-beam actuator with only one beam connecting the anchors at both ends and without a shuttle.


(a)
Displacement/ output force
Hypotenus
α
Opposite
Adjacent
Beams
Central shaft
Anchor
α
Pre-bending angle
L
(b)
Anchor
Displacement/ output force
Beams
θ
(c)
100 μm
Anchors
+
−
+ = To positive terminal of voltage
− = To negative terminal of voltage
Shuttle
(d)

11.21 Schematic images of actuation: (a) Chevron-type horizontal actuator; (b) modified chevron for vertical actuation;[146] (c) top view of a conveyor driven by V-type actuator;[98] (d) contoured beam.[145]


(a)
Secondary unit
PolySi
Anchor
Al
Primary unit
Lift
Si substrate
Si_3N_4
SiO_2
(b)
Secondary unit (no current)
Secondary tip motion
β
w
L_1
2γ
θ
γ
φ
L_2
Primary unit (with current)
Angle bisector of primary unit
(c)

11.22 Cascaded bent-beam actuator: (a) three-dimensional schematic diagram; (b) planar schematic diagram by Zhang *et al.*;[90] (c) an electrothermal microactuator with Z-shaped beam.[34]

11.3.6 Embedded structure-based (silicon skeleton) polymeric electrothermal actuators

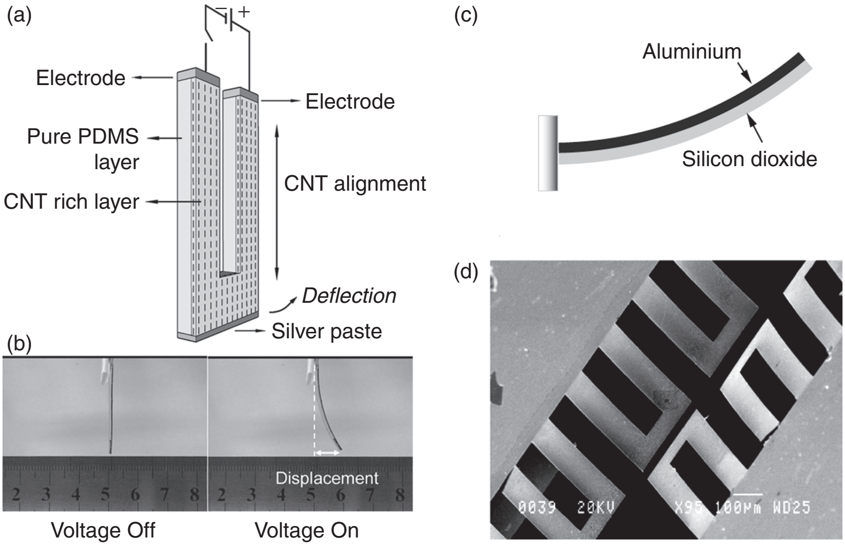
Possessed of unique characteristics,[147,148] the SU-8 polymer is attractive as an electrothermal actuator that offers low handling forces, large displacement and low operating voltages. Therefore, the SU-8 polymer is suitable for the handling and manipulation of micro- or nanoscale and biological applications.[147,149,150] However, their low heat conductivity leads to a slower thermal response.[148,149,151] Lau *et al.* presented a novel concept of embedding silicon structures topped with aluminium track within the SU-8 epoxy with a 30% silicon to 70% SU-8 ratio. This improves the properties of the actuator, such as Young's modulus, thermal stress and work density (by more than 2.5 times), thermal expansion (by 5% more) and response time (by 20%).[152] The aluminium film track devised by Lau *et al.* serves to enable Joule heating by allowing a current to pass through it and subsequently heat up the silicon skeleton underneath. Another configuration fills the gap between the silicon comb fingers with the polymer. Confinement of the polymer inside the high-aspect ratio silicon structure leads to higher displacement, higher stiffness, and less out-of-plane motion. It has been reported that a 2DOF silicon skeleton polymeric microgripper produces 17 μm, 196 μN and 11 μm, 814 μN of displacement and output forces in *x*- and y-directions, respectively, but reaches a maximum of only 250°C at 2.5 V input voltage.[149]

11.3.7 Bimorph actuator

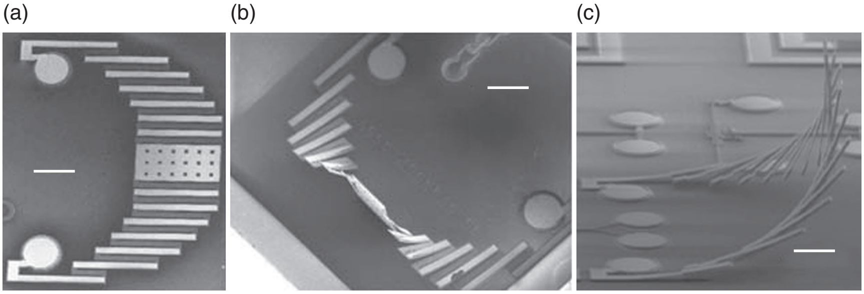
In electrothermal bimorph actuators, mechanical motion is achieved by the differential expansion of two or more different materials, with a large variation in the coefficient of thermal expansion (CTE). Materials with a higher thermal expansion coefficient, such as platinum or gold, usually comprise the top layer.[94] Bimorphs are initially curled up due to the intrinsic stress that is generated in the structural layers during fabrication, but they flatten when voltage is applied.[91,94] Bimorph thermal actuators are advantageous because of their simplicity in fabrication, large array integration, large actuation range and low voltage operation, which makes them suitable for parallel processing and large throughput applications. However, such multilayer structures will suffer from delamination and the top layer will be damaged at elevated temperatures,[153] although bimorphs of titanium tungsten (Ti-W) and silicon nitride (SiN) have been shown to be capable of sustaining large deflection without delamination (Fig. 11.23).[154] Displacement amplification can be carried out by cascading bimorphs.[91,155] Also, rotary and linear actuators have been designed based on this principle (see Fig. 11.24).[155] Recently, a bimorph made of super-aligned carbon nanotube sheets (SACNS) and polydimethylsiloxane (PDMS) composites was demonstrated to be able to reach 9.5 mm at 98°C (with a displacement coefficient of 44 μm/[100 μm × °C]). This bimorph is highly reliable and has a long working life because no delamination was found in it. However, the power consumption can be quite high.[156]

11.3.8 Cantilever type piezoelectric actuators

A piezoelectric cantilever is usually composed of at least one layer of piezoelectric film and one layer of a passive substrate such as a nickel alloy, titanium or silicon, which operates in an out-of-plane motion. A unimorph structure consists of only one layer of piezoelectric film, usually on the top, while bimorph structures are made of two layers of piezoelectric materials that sandwich the centre shim. Sometimes, the piezoelectric film is deposited in the centre of cantilever, coated with passive substrates and electrodes. The characteristics of piezoelectric materials make them good candidates for many applications, such as driving micromirrors and atomic force microscopy (AFM).[157] Also, there have been reports of an oscillating PZT thick film (thickness of 1–50 μm[158]) on a nozzle that changed an outlet pressure and increased a fluid flow

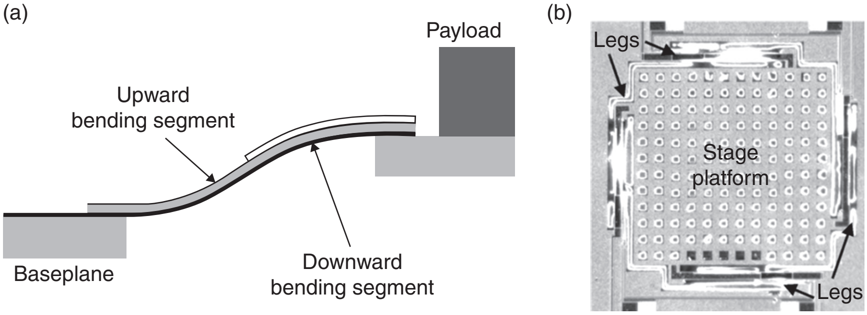

(a)
−
+
Electrode
Electrode
Pure PDMS layer
CNT rich layer
CNT alignment
Deflection
Silver paste
(b)
Displacement
2 3 4 5 6 7 8
2 3 4 5 6 7 8
Voltage Off
Voltage On
(c)
Aluminium
Silicon dioxide
(d)
0039 20KV X95 100Pm WD25

11.23 (a) Schematic structure of the SACNS/PDMS composite actuator. The dashed lines represent the direction of CNT alignment, (b) photographs of an actuator without (left) and with (right) an applied DC voltage of 40 V,[156] (c) schematic of a conventional bimorph cantilever,[23] (d) SEM image of a U-shaped Ti-W/SiN bimorph.[154]



11.24 SEM images of: (a) initial fabricated position of cascaded bimorph scanning mirror; (b) plan view; (c) side view of final position with increased residual stress.[155]

rate.[71] As previously mentioned, small displacement and fabrication incompatibility are major setbacks for piezoelectric actuators. Researchers have sought to create designs that will avoid these problems. Deposition of a gold layer on either end of the PZT beam changes the overall system neutral axis, thus enhancing the displacement in one direction. An actuator 500 μm long and 100 μm wide with 50% gold coverage on the top surface of the actuator (25% of the actuator covered on either end) was reported to be able to apply 7 mN



11.25 Schematic of: (a) bend-up/bend down actuator lifting a load; (b) SEM top view image of vertical stage actuated by bend-up/bend down piezoelectric actuator.[159]

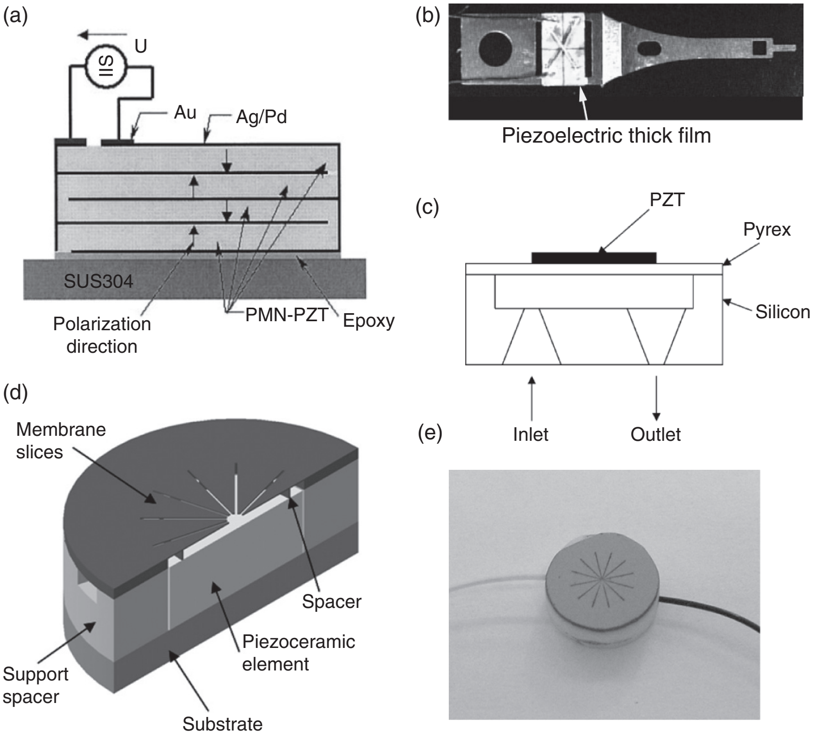
force at almost 1 μm displacement, as well as displacements near 5 μm against a 25 μm opposing force at 20 V.[66] Using a similar approach and concept, a set of four 920 μm × 70 μm compound bend-up/bend-down unimorph microactuators that were used to drive a vertical stage have been shown to achieve displacement as high as 120 μm (see Fig. 11.25).[159] A micromirror with a Wu *et al.* design (actuated by three cascaded electrothermal bimorphs)[91] has been built using piezoelectric thin film (<1 μm[158]) unimorph actuators. Applying 3.5 V_{pp} moved the micromirror 32 μm vertically.[160] Other approaches included adding other mechanics (e.g. introducing an aluminium flexure that connects both ends of the piezoelectric cantilever). When voltage is applied, the piezoelectric cantilever will deform and compress the aluminium cap so that it buckles outwards, giving more displacement.[67] A hybrid thermopiezoelectric microactuator has recently been proposed to extend the displacement. A unimorph composed of PZT and copper is connected to the top of a Peltier module, also known as a 'thermoelectric cooler' (TEC), to give greater variation in the temperature between the two materials, hence increasing the deflection.[161] New fabrication methods and alternative piezoelectric materials have also been proposed. Instead of PZT, AlN has been reported to be a post-CMOS compatible piezoelectric material, since aluminium and nitride are common materials in IC fabrication.[162] However, AlN has a poor piezoelectric constant, giving approximately 15 nm of displacement at 5 V,[162] making it more suitable for nanoscale applications. Andrei *et al.* showed that displacement can be extended by using a thicker AlN layer with non-standard CMOS metals, such as chromium (Cr), as electrodes and a higher excitation voltage.[163]

11.3.9 Membrane type piezoelectric actuators

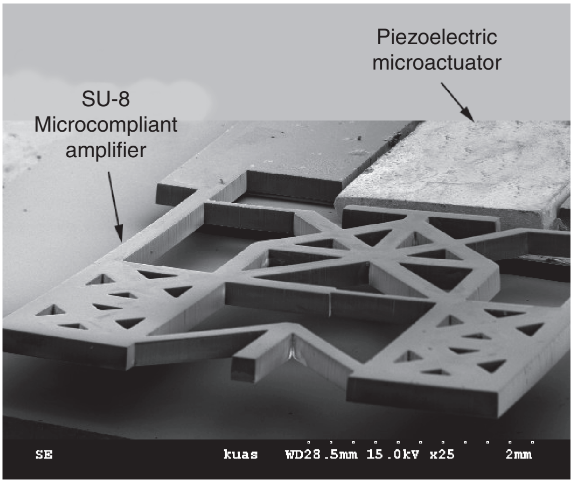
Membrane type piezoelectric actuators can be either thin or thick piezoelectric film, or bulk piezoelectric actuators (50–1 mm[158]) such as stacks or bimorph piezo-discs or films, which have one of their wide area sides completely anchored to either a bulk substrate or a passive membrane. Adhesive epoxy is used to glue these piezoelectric diaphragms to the substrates. Membrane type actuators are used for applications such as micropumps,[164] microvalves,[165] micromotors[166] and hard disk drive head positioning.[13] Their displacements are relatively small at less than 1 μm.[164,166] However, the displacement can be enlarged using a larger sized thermal actuator.[167] Using a stack of piezoelectric diaphragms with IDT electrodes can increase the deflection of a membrane type actuator.[73,168] The IDT diaphragms, which are 800 μm in diameter and 2.8 μm thick, are reported to generate around 7.0 μm centre deflection. Other mechanisms can be added to convert the large force exerted by the PZT into higher displacement via a micro-compliant amplifier, as has been achieved by Huang and Lan.[79] Bolzmacher *et al.* used a silicon membrane leverage unit to amplify the piezoelectric actuator displacement. The piezoelectric bulk actuator is placed in a slot at the centre of the unit and covered by membrane slices. When voltage is applied, the piezoelectric actuator will push the membrane slice upwards, where the tip will reach its highest position (see Fig. 11.26d and 11.26e).[169]

11.3.10 Compliant mechanism

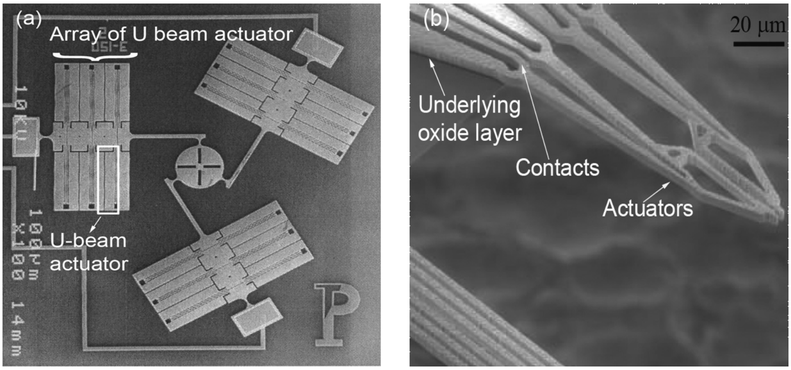
Compliant mechanisms are monolithic structures that utilize their flexible structures in order to transmit motion or force from an actuator.[170] For example, a micro-compliant made of SU-8 (Fig. 11.27) with an average amplifier factor of 6.9 has successfully produced approximately 15.2 μm and 730 μN at 20 V from a piezoelectric actuator.[79] Large forces from the piezoelectric actuator have been converted into a greater displacement. In electrothermal compliant mechanisms (ETCs), elastic deformation due to thermal expansion provides the actuation force to the revolute joints hinges and flexures and amplifies the motion. There have been demonstrations of various ETC mechanisms that use a combination of U-beam actuators to achieve lateral translation and expansion.[135] Parallel configuration is used in 3DOF planar micromanipulators. Tsai *et al.*[171] proposed 28 different feasible configurations to transform the conventional rigid links kinematic macromechanism into ETC mechanisms. Compliant designs have been developed by topology optimization,[79,172] which use a finite element-based method with a multiparameter optimization algorithm for the optimal material distribution within a fixed domain. Compared with a pair of three-beam designs[33] (Fig. 11.20c) without topology optimization, a topology-optimized electrothermal actuator (see Fig. 11.28b)[172] is able to provide a higher gripping force and lower end-effector temperature within the same


(a)
U
Au
Ag/Pd
SUS304
Polarization direction
PMN-PZT
Epoxy
(b)
Piezoelectric thick film
(c)
PZT
Pyrex
Silicon
Inlet
Outlet
(d)
Membrane slices
Spacer
Piezoceramic element
Support spacer
Substrate
(e)

11.26 (a) Cross-sectional view of a piezoelectric actuator with four layers;[73] (b) photograph of a piezoelectric quadmorph-actuated hard disk drive suspension;[13] (c) cross-sectional view of a valveless micropump;[164] (d) 3D schematic of a piezoelectric actuator with leverage unit; (e) photograph of a piezoelectric actuator with leverage unit.[169]


Piezoelectric microactuator
SU-8 Microcompliant amplifier
SE kuas WD28.5mm 15.0kV x25 2mm

11.27 SEM image of a piezoelectric microactuator with micro-compliant amplifier.[79]



11.28 SEM images: (a) 3DOF planar micro-manipulator with array of U-beam actuators in parallel configuration;[136] (b) topology-optimized electrothermal actuator microgripper.[172]

geometric domain as the double-U-beam microactuator. The end-effector was measured to be at 229°C and deflection was up to 1 μm with a gripping force of 18–20 μN.

11.4 Conclusion

Microactuators have shown a promising future and are important in many fields as the demand continues for highly functional items at a reduced size. This is the case for nanoscale applications, which require tools and processes that are capable of precisely positioning and manipulating nanoscale components and materials. This chapter reviewed the state-of-the-art in microactuators and related issues, examining design, fabrication and integration with ICs. Basic working principles, design considerations, the selection of different transduction mechanisms for operation and design optimization were discussed.

With the relentless increase in enabling technologies, the MEMS foundry services offered to designers will continue to expand. However, the difficulty of matching the current and future needs of the MEMS designer with the limitations of commercial MEMS foundry processes should not be underestimated. For example, many designs that have been demonstrated were custom-made in laboratories especially for piezoelectric microactuators and they are not yet supported by commercial foundries. This has impeded the growth and employment of piezoelectric microactuators, especially in high-speed, low-power applications, although this type of microactuator has high capabilities.

In order to gain enhanced performances and compactness, the integration of MEMS with CMOS has become one of the main research interests in this field. Most of the micromachining processes available have been shown to be compatible with standard IC technologies. A considerable number of breakthroughs have been achieved in this matter. Shortcomings of a particular technique were identified and a solution was found to overcome them. Again, it is the piezoelectric microactuators that need further research and investigation. Although alternative materials (e.g. AlN) have been found to be fully compatible with current CMOS fabrication processes, piezoelectric capability needs to be improved.

11.5 References

1. M. Gad-el-Hak (2002), 'Introduction', in M. Gad-el-Hak (Ed.), *The MEMS Handbook*, ch 1, CRC Press, Boca Raton, Florida, pp. 1–5.
2. E. Thielicke and E. Obermeier (June 2000), 'Microactuators and their technologies', *Mechatronics*, vol. **10**, no. 4–5, pp. 431–55.
3. F. Laermer (2005), 'Mechanical microsensors', in J. Korvink and O. Haber (Eds.), *MEMS: A Practical Guide to Design, Analysis, and Applications*, Norwich, New York: William Andrew Publishing.
4. S. A. Vittorio (October 2001), 'Micro electro mechanical systems (MEMS)', *Cambridge Scientific Abstracts* [Online]. Available at: http://www.csa.com/discoveryguides/mems/overview.php.
5. H. Fujita (August 1998), 'Microactuators and micromachines', *Proceedings of the IEEE*, vol. **86**, no. 8, pp. 1721–32.
6. B. Sahu, C. R. Taylor and K. K. Leang (June 2010), 'Emerging challenges of microactuators for nanoscale positioning, assembly, and manipulation', *Journal of Manufacturing Science and Engineering*, vol. **132**, no. 3, pp. 030917 (16pp).
7. J. W. Judy (2005), 'Microactuators', in J. Korvink and O. Haber (Eds.), *MEMS: A Practical Guide to Design, Analysis, and Applications*, William Andrew Publishing, Norwich, New York.
8. N. Tirole, D. Hauden, P. Blind, M. Froelicher and L. Gaudriot (May 1995), 'Three-dimensional silicon electrostatic linear micro actuator', *Sensors and Actuators A: Physical*, vol. **48**, no. 2, pp. 145–50.
9. M. Komori and T. Hirakawa (June 2005), 'A magnetically driven linear microactuator with new driving method', *Transactions on Mechatronics*, *IEEE/ASME*, vol. **10**, no. 3, pp. 335–8.
10. P. Gao, K. Yao, X. Tang, X. He, S. Shannigrahi, Y. Lou, J. Zhang and K. Okada (August 2006), 'A piezoelectric micro-actuator with a three-dimensional structure and its micro-fabrication', *Sensors and Actuators A: Physical*, vols **130–1**, pp. 491–6.
11. L. Li and D. Uttamchandani (June 2009), 'Dynamic response modelling and characterization of a vertical electrothermal actuator', *Journal of Micromechanics and Microengineering*, vol. **19**, no. 7, pp. 075014 (9pp).
12. Z. W. Zhong and J. Sun (June 2004), 'A suspension integrated with a piezoelectric microactuator for dual stage actuation', *International Journal of Advanced Manufacturing Technology*, vol. **24**, no. 9–10, pp. 686–92.
13. G. K. Lau and H. Du (July 2006), 'A piezoelectric quad-morph actuated suspension for hard disk drives', *Journal of Micromechanics and Microengineering*, vol. **16**, no. 9, pp. 1833–40.

14. J. Chen, Y. Lu and J. P. Yang (May 2005), 'Integrated fabrication of electrostatic microactuator for HDD R/W head positioning', *Microsystem Technologies*, vol. **11**, no. 8–10, pp. 612–15.
15. H. Toshiyoshi, M. Mita and H. Fujita (December 2002), 'A MEMS piggyback actuator for hard-disk drives', *Journal of Microelectromechanical Systems*, vol. **11**, no. 6, pp. 648–54.
16. D. Dinulovic, H. Saalfeld, Z. Celinski, S. B. Field and H. H. Gatzen (November 2008), 'Integrated electromagnetic second stage microactuator for a hard disk recording head', *IEEE Transactions on Magnetics*, vol. **44**, no. 11, pp. 3730–3.
17. G. Zhou, J. Chen and D. Ning (2010), 'A novel electromagnetic microactuator for dual-stage position of hard disk', in *International Conference on E-Product E-Service and E-Entertainment (ICEEE)*, Henan, 7–9 November 2010, pp. 1–4.
18. L. Li (August 2010), 'Two-dimensional moveable optical pinhole with applications in single-detector imaging systems', *Electronics Letters*, vol. **46**, no. 17, pp. 1215–17.
19. Z. Cui, X. Wang, Y. Li and G. Y. Tian (April 2007), 'High sensitive magnetically actuated micromirrors for magnetic field measurement', *Sensors and Actuators A: Physical*, vol. **138**, no. 1, pp. 145–50.
20. C. -H. Ji and Y. -K. Kim (May 2003), 'Electromagnetic micromirror array with single-crystal silicon mirror plate and aluminum spring', *Journal of Lightwave Technology*, vol. **21**, no. 3, pp. 584–90.
21. Y. Okano and Y. Hirabayashi (January/February 2002), 'Magnetically actuated micromirror and measurement system for motion characteristics using specular reflection', *IEEE Journal of Selected Topics in Quantum Electronics*, vol. **8**, no. 1, pp. 19–25.
22. O. Cugat, S. Basrour, C. Divoux, P. Mounaix and G. Reyne (March 2001), 'Deformable magnetic mirror for adaptive optics: technological aspects', *Sensors and Actuators A: Physical*, vol. **89**, no. 1–2, pp. 1–9.
23. S. Pal and H. Xie (July 2010), 'Pre-shaped open loop drive of electrothermal micromirror by continuous and pulse width modulated waveforms', *IEEE Journal of Quantum Electronics*, vol. **46**, no. 9, pp. 1254–60.
24. S. Pal, K. Jia and H. Xie (2007), 'An electrothermal micromirror with high linear scanning efficiency', in *The 20th Annual Meeting of the IEEE Lasers and Electro-Optics Society, 2007. LEOS 2007.*, Lake Buena Vista, FL, 21–25 October 2007, pp. 914–15.
25. T. Bakke, A. Vogl, O. Żero, F. Tyholdt, I. -R. Johansen and D. Wang (June 2010), 'A novel ultra-planar, long-stroke and low-voltage piezoelectric micromirror', *Journal of Micromechanics and Microengineering*, vol. **20**, no. 6, pp. 064010 (7pp).
26. Y. -H. Jin, K. -S. Seo, Y. -H. Cho, S. -S. Lee, K. -C. Song and J. -U. Bu (September 2004), 'An optical microswitch chip integrated with silicon waveguides and touch-down electrostatic micromirrors', *Journal of Micromechanics and Microengineering*, vol. **14**, no. 12, pp. 1674–81.
27. J. Kim, D. Christensen and L. Lin (November 2005), 'Monolithic 2-D scanning mirror using self-aligned angular vertical comb drives', *IEEE Photonics Technology Letters*, vol. **17**, no. 11, pp. 2307–9.
28. R. Sulima and S. Wiak (2008), 'Modelling of vertical electrostatic comb-drive for scanning micromirrors', *International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, vol. **27**, no. 4, pp. 780–7.

29. W. Trimmer and R. H. Stroud (2002), 'Scaling of micromechanical devices', in M. Gad-el-Hak (Ed.), *The MEMS Handbook*, ch. 2, CRC Press, Boca Raton, Florida, pp. 6–18.
30. A. M. Petrina and A. A. Petrin (December 2007), 'On the question of information processing in nanomanipulation tasks' *Computer Science*, vol. **42**, no. 2, pp. 136–42.
31. D. Sinan Haliyo, S. Regnier and J. -C. Guinot (November/December 2003), ' [mü] MAD, the adhesion based dynamic micro-manipulator', *European Journal of Mechanics – A/Solids*, vol. **22**, no. 6, pp. 903–16.
32. W. Benecke (1991), 'Silicon-microactuators: activation mechanisms and scaling problems', in *International Conference on Solid-State Sensors and Actuators, 1991. Digest of Technical Papers, TRANSDUCERS '91*, Francisco, CA, USA, pp. 46–50.
33. K. N. Andersen, D. H. Petersen, K. Carlson, K. Molhave, O. Sardan, A. Horsewell, V. Eichhorn, S. Fatikow and P. Boggild (January 2009), 'Multimodal electrothermal silicon microgrippers for nanotube manipulation', *IEEE Transactions Nanotechnology*, vol. **8**, no. 1, pp. 76–85.
34. C. Guan and Y. Zhu (July 2010), 'An electrothermal microactuator with Z-shaped beams', *Journal of Micromechanics and Microengineering*, vol. **20**, no. 8, pp. 085014 (9pp).
35. C. Elbuken, N. Topaloglu, P. M. Nieva, M. Yavuz and J. P. Huissoon (February 2009), 'Modeling and analysis of a 2-DOF bidirectional electro-thermal microactuator', *Microsystem Technologies*, vol. **15**, no. 5, pp. 713–22.
36. J. Singh, A. Agarwa and M. Soundarapandian (May 2006), 'A novel electrostatic microactuator for large deflections in MEMS applications', *Thin Solid Films*, vol. **504**, no. 1–2, pp. 64–8.
37. S. C. Jacobsen, R. H. Price, J. E. Wood, T. H. Rytting and M. Rafaelof (1989), 'A design overview of an eccentric-motion electrostatic microactuator (the wobble motor)', *Sensors and Actuators*, vol. **20**, no. 1–2, pp. 1–16.
38. A. Jungen, M. Pfenninger, M. Tonteling, C. Stampfer and C. Hierold (July 2006), 'Electrothermal effects at the microscale and their consequences on system design', *Journal of Micromechanics and Microengineering*, vol. **16**, no. 8, pp. 1633–8.
39. O. Ozsun, B. Erdem Alaca, A. D. Yalcinkaya, M. Yilmaz, M. Zervas and Y. Leblebici (March 2009), 'On heat transfer at microscale with implications for microactuator design', *Journal of Micromechanics and Microengineering*, vol. **19**, no. 4, pp. 045020 (13pp).
40. E.S. Kolesar, M. D. Ruff, W. E. Odom, J. A. Jayachadran, J. B. McAllister, S. Y. Ko, J. T. Howard, P. B. Allen, J. M. Wilken, N. C. Boydston, J. E. Bosch and R. J. Wilks (December 2002), 'Single- and double-hot arm asymmetrical polysilicon surface micromachined electrothermal microactuators applied to realize a microengine', *Thin Solid Films*, vol. **420–1**, no. 2, pp. 530–8.
41. V. A. Lisovsky and S. D. Yakovin (June 2000), 'Scaling law for a low-pressure gas breakdown in a homogeneous DC electric field', *JETP Letters*, vol. **72**, no. 2, pp. 34–7.
42. K. Kim, X. Liu, Y. Zhang and Y. Sun (April 2008), 'Nanonewton force-controlled manipulation of biological cells using a monolithic MEMS microgripper with two-axis force feedback', *Journal of Micromechanics and Microengineering*, vol. **18**, no. 5, pp. 055013 (8pp).

43. Yao Fu, M. K. Ghantasala, E. Harvey and L. Qin (May 2005), 'Design and fabrication of a hybrid actuator', *Smart Materials and Structures*, vol. **14**, no. 4, pp. 488–95.
44. M. Kiuchi, S. Matsui and Y. Isono (April 2008), 'Mechanical characteristics of FIB deposited carbon nanowires using an electrostatic actuated nano tensile testing device', *Journal of Microelectromechanical Systems*, vol. **16**, no. 2, pp. 191–201.
45. M. Mita, M. Arai, S. Tensaka, D. Kobayashi and H. Fujita (February 2003), 'A micromachined impact microactuator driven by electrostatic force', *Journal of Microelectromechanical Systems*, vol. **12**, no. 1, pp. 37–41.
46. H. Zhu, J. Miao, B. Chen, Z. Wang and W. Zhu (May 2005), 'Membrane microcantilever arrays fabrication with PZT thin films', *Microsystem Technologies*, vol. **11**, no. 8–10, pp. 1121–6.
47. W. Dong, X. Lu, M. Liu, Y. Cui and J. Wang (January 2007), 'Measurement on the actuating and sensing capability of a PZT microcantilever', *Measurement Science and Technology*, vol. **18**, no. 3, pp. 601–8.
48. C. Huang, C. Christophorou, K. Najafi, A. Naguib and H. M. Nagib (June 2002), 'An electrostatic microactuator system for application in high-speed jets', *Journal of Microelectromechanical Systems*, vol. **11**, no. 3, pp. 222–35.
49. S. Fu, G. Ding, Hong Wang, Zhuoqing Yang and Jianzhi Feng (April 2007), 'Design and fabrication of a magnetic bi-stable electromagnetic MEMS relay', *Microelectronics Journal*, vol. **38**, no. 4–5, pp. 556–63.
50. V. T. Srikar and S. M. Spearing (January 2003), 'Materials selection for microfabricated electrostatic actuators', *Sensors and Actuators A: Physical*, vol. **102**, no. 3, pp. 279–85.
51. I. –J. Cho, T. Song, S. -H. Baek and E. Yoon (July 2005), 'A low-voltage and low-power RF MEMS series and shunt switches actuated by combination of electromagnetic and electrostatic forces', *IEEE Transactions on Microwave Theory and Techniques*, vol. **53**, no. 7, pp. 2450–7.
52. Y. W. Huang, T. -Y. Chao, C. C. Chen and Y. T. Cheng (June 2007), 'Power consumption reduction scheme of magnetic microactuation using electroplated Cu–Ni nanocomposite', *Applied Physics Letters*, vol. **90**, no. 24, pp. 244105 (3pp).
53. L. Li and D. Uttamchandani (September 2004), 'Modified asymmetric micro-electrothermal actuator: analysis and experimentation', *Journal of Micromechanics and Microengineering*, vol. **14**, no. 12, pp. 1734–41.
54. W. Dai, K. Lian and W. Wang (February 2007), 'Design and fabrication of a SU-8 based electrostatic microactuator', *Microsystem Technologies*, vol. **13**, no. 3–4, pp. 271–7.
55. P. Cheung, R. Horowitz and R. T. Rowe (January 1996), 'Design, fabrication, position sensing, and control of an electrostatically-driven polysilicon microactuator', *IEEE Transactions on Magnetics*, vol. **32**, no. 1, pp. 122–8.
56. K. Takahashi, E. Bulgan, Y. Kanamori and K. Hane (April 2009), 'Submicrometer comb-drive actuators fabricated on thin single crystalline silicon layer', *IEEE Transactions on Industrial Electronics*, vol. **56**, no. 4, pp. 991–5.
57. P. Srinivasan, C. O. Gollasch and M. Kraft (June 2010), 'Three dimensional electrostatic actuators for tunable optical micro cavities', *Sensors and Actuators A: Physical*, vol. **161**, no. 1–2, pp. 191–8.

58. Yu Sun, D. Piyabongkarn, A. Sezen, B. J. Nelson and R. Rajamani (December 2002), 'A high-aspect-ratio two-axis electrostatic microactuator with extended travel range', *Sensors and Actuators A: Physical*, vol. **102**, no. 1–2, pp. 49–60.
59. H. N. Kwon, I. -H. Hwang and J. -H. Lee (June 2005), 'A pulse-operating electrostatic microactuator for bi-stable latching', *Journal of Micromechanics and Microengineering*, vol. **15**, no. 8, pp. 1511–16.
60. K. B. Lee and Y. -H. Cho (March 2001), 'Laterally driven electrostatic repulsive-force microactuators using asymmetric field distribution', *Journal of Microelectromechanical Systems*, vol. **10**, no. 1, pp. 128–36.
61. T. -S. Kim and H. C. Lee (September 2006), 'A highly sensitive bolometer structure with an electrostatic-actuated signal bridge', *IEEE Transactions on Electron Devices*, vol. **53**, no. 9, pp. 2392–400.
62. A. Koga, K. Suzumori, H. Sudo, S. Iikura and M. Kimura (January 1999), 'Electrostatic linear microactuator mechanism for focusing a CCD camera', *Journal of Lightwave Technology*, vol. **17**, no. 1, pp. 43–7.
63. K. B. Lee (September 2007), 'Non-contact electrostatic microactuator using slit structures: theory and a preliminary test', *Journal of Micromechanics and Microengineering*, vol. **17**, no. 11, pp. 2186–96.
64. C. -L. Dai (May 2003), 'In situ electrostatic microactuators for measuring the Young's modulus of CMOS thin films', *Journal of Micromechanics and Microengineering*, vol. **13**, no. 5, pp. 563–7.
65. I. A. Ivan, M. Rakotondrabe, P. Lutz and N. Chaillet (June 2009), 'Quasistatic displacement self-sensing method for cantilevered piezoelectric actuators', *Review of Scientific Instruments*, vol. **80**, no. 6, pp. 065102 (8pp).
66. K. R. Oldham, J. S. Pulskamp, R. G. Polcawich and M. Dubey (August 2008), 'Thin-film PZT lateral actuators with extended stroke', *Journal of Microelectromechanical Systems*, vol. **17**, no. 4, pp. 890–9.
67. H. K. R. Kommepalli, H. G. Yu, C. L. Muhlstein, S. Trolier-McKinstry, C. D. Rahn and S. A. Tadigadapa (June 2009), 'Design, fabrication and performance of a piezoelectric uniflex microactuator', *Journal of Microelectromechanical Systems*, vol. **18**, no. 3, pp. 616–25.
68. Z. Wang, W. Zhu and X. Yao (October 2002), 'd31 Type inplane bending multilayer piezoelectric microactuators – a design concept and its applications', *Sensors and Actuators A: Physical*, vol. **101**, no. 3, pp. 262–8.
69. T. Morita, K. Shimizu, M. Hasegawa, K. Oka and T. Higuchi (September 2002), 'A miniaturized levitation system with motion control using a piezoelectric actuator', *IEEE Transactions on Control Systems Technology*, vol. **10**, no. 5, pp. 666–670.
70. K. Tsuchiya, T. Kitagawa and E. Nakamachi (July 2003), 'Development of RF magnetron sputtering method to fabricate PZT thin film actuator', *Precision Engineering*, vol. **27**, no. 3, pp. 258–64.
71. R. S. Wiederkehr, M. C. Salvadori, J. Brugger, F. T. Degasperi and M. Cattani (May 2008), 'The gas flow rate increase obtained by an oscillating piezoelectric actuator on a micronozzle', *Sensors and Actuators A: Physical*, vol. **144**, no. 1, pp. 154–60.
72. S. Zurn, M. Hsieh, G. Smith, D. Markus, M. Zang, G. Hughes, Y. Nam, M. Arik and D. Polla *et al.*, 'Fabrication and structural characterization of a resonant frequency PZT microcantilever', *Smart Materials and Structures*, vol. **10**, no. 2, pp. 252–63, January 2001.

73. Y. Jing and J. Luo (May 2005), 'Structure and electrical properties of PMN-PZT micro-actuator deposited by tape-casting process', *Journal of Materials Science: Materials in Electronics*, vol. **16**, no. 5, pp. 287–94.
74. K. -M. Chang, R. -J. Lin and I. -C. Deng (March 2007), 'Design of low-temperature CMOS-process compatible membrane fabricated with sacrificial aluminum layer for thermally isolated applications', *Sensors and Actuators A: Physical*, vol. **134**, no. 2, pp. 660–7.
75. T. Tamagawa, D. L. Polla and C. C. Hsueh (1990), 'Lead zirconate titanate (PZT) thin films in surface-micromachined sensor structures', in *Technical Digest., International Electron Devices Meeting, 1990. IEDM '90.*, San Francisco, CA, USA, 9–12 December 1990, pp. 617–20.
76. S. Trolier-McKinstry and P. Muralt (January 2004), 'Thin film piezoelectrics for MEMS', *Chemistry and Materials Science*, vol. **12**, no. 1–2, pp. 7–17.
77. M. Ferrari, V. Ferrari, M. Guizzetti and D. Marioli (2009), 'Piezoelectric low-curing-temperature ink for sensors and power harvesting', in Piero Malcovati, A. Baschirotto, A. d'Amico and C. Natale (Eds.), *Sensors and Microsystems: AISEM 2009 Proceedings*, Springer, Pavia, Italy, 24–26 February 2009, pp. 77–81.
78. C. Giordano, I. Ingrosso, M. T. Todaro, G. Maruccio, S. De Guido, R. Cingolani, A. Passaseo and M. De Vittorio (April 2009), 'AlN on polysilicon piezoelectric cantilevers for sensors/actuators', *Microelectronic Engineering*, vol. **86**, no. 4–6, pp. 1204–7.
79. S. -C. Huang and G. -J. Lan (February 2006), 'Design and fabrication of a micro-compliant amplifier with a topology optimal compliant mechanism integrated with a piezoelectric microactuator', *Journal of Micromechanics and Microengineering*, vol. **16**, no. 3, pp. 531–8.
80. E. Sicard and S. D. Bendhia (2007), *Basics of CMOS Cell Design*. Toulouse, France, McGraw-Hill.
81. E. Sicard and S. D. Bendhia (2007), *Advanced CMOS Cell Design*. Toulouse, France: McGraw-Hill.
82. M. -N. Sabry (March 2003), 'Compact thermal models for electronic systems', *IEEE Transactions on Components and Packaging Technologies*, vol. **26**, no. 1, pp. 179–85.
83. M. -N. Sabry (November 1999), 'Static and dynamic thermal modeling of ICs', *Microelectronics Journal*, vol. **30**, no. 11, pp. 1085–91.
84. F. Mohammadi and M. Marami (2008), 'Creation and verification of dynamic compact thermal model of a BGA package', *Microelectronics International*, vol. **25**, no. 3, pp. 3–13.
85. Y. Shabany (2002), 'Component size and effective thermal conductivity of printed circuit boards', in *The Eighth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, 2002. ITHERM 2002*, San Diego, California, 30 May–1 June, pp. 489–94.
86. B. S. H. Royce (December 1988), 'Differential thermal expansion in microelectronic systems', *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, vol. **11**, no. 4, pp. 454–63.
87. T. Bechtold, E. B. Rudnyi and J. G. Korvink (October 2005), 'Dynamic electro-thermal simulation of microsystems – a review', *Journal of Micromechanics and Microengineering*, vol. **15**, no. 11, pp. R17–R31.
88. T. H. K Barron (1998), 'Generalized theory of thermal expansion of solids', in *Thermal Expansion of Solids*. ch. 1, ASM International, pp. 1–105.

89. A. Cao, J. Kim and L. Lin (April 2007), 'Bi-directional electrothermal electromagnetic actuators', *Journal of Micromechanics and Microengineering*, vol. **17**, no. 5, pp. 975–82.
90. Y. Zhang, Q. -A. Huang, R. -G. Li and W. Li (March 2006), 'Macro-modeling for polysilicon cascaded bent beam electrothermal microactuators', *Sensors and Actuators A: Physical*, vol. **128**, no. 1, pp. 165–75.
91. L. Wu and H. Xie (July–August 2008), 'A large vertical displacement electrothermal bimorph microactuator with very small lateral shift', *Sensors and Actuators A: Physical*, vols **145–6**, pp. 371–9.
92. E.S. Kolesar, S. Y. Ko, J. T. Howard, P. B. Allen, J. M. Wilken, N. C. Boydston, M. D. Ruff and R. J. Wilks (December 2000), 'In-plane tip deflection and force achieved with asymmetrical polysilicon electrothermal microactuators', *Thin Solid Films*, vols **377–8**, pp. 719–26.
93. R. R. A. Syms, H. Zou and J. Stagg (May 2006), 'Micro-opto-electro-mechanical systems alignment stages with Vernier latch mechanisms', *Journal of Optics A: Pure and Applied Optics*, vol. **8**, no. 7, pp. S305–S312.
94. M. Y. Al Aioubi, V. Djakov, S. E. Huq and P. D. Prewett (June 2004), 'Deflection and load characterisation of bimorph actuators for bioMEMS and other applications', *Microelectronic Engineering*, vol. **73–74**, pp. 898–903.
95. Long Que, Jae-Sung Park and Y. B. Gianchandani (June 2001), 'Bent-beam electrothermal actuators-Part I: Single beam and cascaded devices', *Journal of Microelectromechanical Systems*, vol. **10**, no. 2, pp. 247–54.
96. R. S. Chen, C. Kung and G. -B. Lee (April 2002), 'Analysis of the optimal dimension on the electrothermal microactuator', *Journal of Micromechanics and Microengineering*, vol. **12**, no. 3, pp. 291–6.
97. J. W. Judy, T. Tamagawa and D. L. Polla (1990), 'Surface micromachined linear thermal microactuator', in *International Electron Devices Meeting, 1990. IEDM '90. Technical Digest.*, San Francisco, CA, USA, pp. 629–32.
98. B. Shay, T. Hubbard and M. Kujath (November 2008), 'Linear frictional microconveyors', *Sensors and Actuators A: Physical*, vol. **148**, no. 1, pp. 290–8.
99. Lijie Li, M. Begbie, G. Brown and D. Uttamchandani (August 2007), 'Design, simulation and characterization of a MEMS optical scanner', *Journal of Micromechanics and Microengineering*, vol. **17**, no. 9, pp. 1781–7.
100. J. H. Lau, C. K. Lee, C. S. Premachandran and Y. Aibin (2010), 'Optical MEMS packaging: communications', in *Advance MEMS Packaging*. ch. 5, McGraw-Hill, New York, pp. 212–15.
101. A. P. Lee, C. F. McConaghy, G. Sommargren, P. Krulevitch and E. W. Campbell (December 2003), 'Vertical-actuated electrostatic comb-drive with in situ capacitive position correction for application in phase shifting diffraction interferometry', *Journal of Microelectromechanical Systems*, vol. **12**, no. 6, pp. 960–71.
102. D. Elata and V. Leus (May 2005), 'How slender can comb-drive fingers be?', *Journal of Micromechanics and Microengineering*, vol. **15**, no. 5, pp. 1055–9.
103. G. Zhou and P. Dowd (March 2003), 'Tilted folded-beam suspension for extending the stable travel range of comb-drive actuators', *Journal of Micromechanics and Microengineering*, vol. **13**, no. 2, pp. 178–83.
104. T. Hirano, T. Furuhata, K. J. Gabriel and H. Fujita (March 1992), 'Design, fabrication, and operation of submicron gap comb-drive microactuators', *Journal of Microelectromechanical Systems*, vol. **1**, no. 1, pp. 52–9.

105. D. Piyabongkarn, Y. Sun, R. Rajamani, A. Sezen and B. J. Nelson (January 2005), 'Travel range extension of a MEMS electrostatic microactuator', *IEEE Transactions on Control Systems Technology*, vol. **13**, no. 1, pp. 138–45.
106. J. I. Seeger and B. E. Boser (1999), 'Dynamics and control of parallel-plate actuators beyond the electrostatic instability', in *Transducers '99, The 10th International Conference on Solid-State Sensors and Actuators*, Sendai, Japan, pp. 474–7.
107. V. P. Jaecklin, C. Linder, N. F. de Rooij and J. M. Moret (December 1992), 'Micromechanical comb actuators with low driving voltage', *Journal of Micromechanics and Microengineering*, vol. **2**, no. 4, pp. 250–5.
108. W. C. Tang, M. G. Lim and R. T. Howe (December 1992), 'Electrostatic comb-drive levitation and control method', *Journal of Microelectromechanical Systems*, vol. **1**, no. 4, pp. 170–8.
109. M. T. -K. Hou, G. K. -W. Huang, J. -Y. Huang, K. -M. Liao, R. Chen and J. -L. Andrew Yeh (April 2006), 'Extending displacements of comb-drive actuators by adding secondary comb electrodes', *Journal of Micromechanics and Microengineering*, vol. **16**, no. 4, pp. 684–91.
110. R. Legtenberg, A. W. Groeneveld and M. Elwenspoek (September 1996), 'Comb-drive actuators for large displacements', *Journal of Micromechanics and Microengineering*, vol. **6**, no. 3, pp. 320–9.
111. J. D. Grade, H. Jerman and T. W. Kenny (June 2003), 'Design of large deflection electrostatic actuators', *Journal of Microelectromechanical Systems*, vol. **12**, no. 3, pp. 335–43.
112. C. Chen and C. Lee (September 2004), 'Design and modeling for comb-drive actuator with enlarged static displacement', *Sensors and Actuators A: Physical*, vol. **115**, nos 2–3, pp. 530–9.
113. J. -C. Chiou, Y. -J. Lin and C. -F. Kuo (January 2008), 'Extending the traveling range with a cascade electrostatic comb-drive actuator', *Journal of Micromechanics and Microengineering*, vol. **18**, no. 1, pp. 015018 (7pp).
114. B. Borovic, F. L. Lewis, A. Q. Liu, E. S. Kolesar and D. Popa (July 2006), 'The lateral instability problem in electrostatic comb-drive actuators: modeling and feedback control', *Journal of Micromechanics and Microengineering*, vol. **16**, no. 7, pp. 1233–41.
115. E. S. Hung and S. D. Senturia (December 1999), 'Extending the travel range of analog-tuned electrostatic actuators', *Journal of Microelectromechanical Systems*, vol. **8**, no. 4, pp. 497–505.
116. D. -Y. Qiao, W. -Z. Yuan and X. -Y. Li (April 2007), 'A two-beam method for extending the working range of electrostatic parallel-plate micro-actuators', *Journal of Electrostatics*, vol. **65**, no. 4, pp. 256–62.
117. E. K. Chan and R. W. Dutton (September 2000), 'Electrostatic micromechanical actuator with extended range of travel', *Journal of Microelectromechanical Systems*, vol. **9**, no. 3, pp. 321–8.
118. J. I. Seeger and B. E. Boser, 'Charge control of parallel-plate, electrostatic actuators and the tip-in instability', *Journal of Microelectromechanical Systems*, vol. **12**, no. 5, pp. 656–71, October 2003.
119. R. Nadal-Guardia, A. Dehe, R. Aigner and L. M. Castaner, 'Current drive methods to extend the range of travel of electrostatic microactuators beyond the voltage pull-in point', *Journal of Microelectromechanical Systems*, vol. **11**, no. 3, pp. 255–63, June 2002.

120. L. Castañer, J. Pons, R. Nadal-Guardia and A. Rodríguez (May 2001), 'Analysis of the extended operation range of electrostatic actuators by current-pulse drive', *Sensors and Actuators A: Physical*, vol. **90**, no. 3, pp. 181–90.
121. L. A. Rocha, E. Cretu and R. F. Wolffenbuttel (January 2006), 'Using dynamic voltage drive in a parallel-plate electrostatic actuator for full-gap travel range and positioning', *Journal of Microelectromechanical Systems*, vol. **15**, no. 1, pp. 69–83.
122. G. He and Z. Geng (February 2012), 'Finite-time stabilization of a comb-drive electrostatic microactuator', *IEEE/ASME Transactions on Mechatronics*, vol. **17**, no. 1, pp. 107–15.
123. J. Dong, D. Mukhopadhyay and P. M. Ferreira (June 2007), 'Design, fabrication and testing of a silicon-on-insulator (SOI) MEMS parallel kinematics XY stage', *Journal of Micromechanics and Microengineering*, vol. **17**, no. 6, pp. 1154–61.
124. X. Liu, J. Tong and Y. Sun (October 2007), 'A millimeter-sized nanomanipulator with sub-nanometer positioning resolution and large force output', *Smart Materials and Structures*, vol. **16**, no. 5, pp. 1742–50.
125. L. Sun, J. Wang, W. Rong, X. Li and H. Bao (December 2008), 'A silicon integrated micro nano-positioning XY-stage for nano-manipulation', *Journal of Micromechanics and Microengineering*, vol. **18**, no. 12, pp. 125004 (9pp).
126. X. Liu, K. Kim and Y. Sun (September 2007), 'A MEMS stage for 3-axis nanopositioning', *Journal of Micromechanics and Microengineering*, vol. **17**, no. 9, pp. 1796–1802.
127. J. -C. Tsai, S. -J. Chiou, T. -L. Hsieh, C. -W. Sun, D. Hah and M. C. Wu (April 2008), 'Two-axis MEMS scanners with radial vertical combdrive actuators – design, theoretical analysis, and fabrication', *Journal of Optics A: Pure and Applied Optics*, vol. **10**, no. 4, pp. 044006 (7pp).
128. K. P. Tee, S. S. Ge and F. E. H. Tay (March 2009), 'Adaptive control of electrostatic microactuators with bidirectional drive', *IEEE Transactions on Control Systems Technology*, vol. **17**, no. 2, pp. 340–52.
129. R. Legtenberg, J. Gilbert, S. D. Senturia and M. Elwenspoek (September 1997), 'Electrostatic curved electrode actuators', *Journal of Microelectromechanical Systems*, vol. **6**, no. 3, pp. 257–65.
130. C. -Y. Shao, Y. Kawai, M. Esashi and T. Ono (August 2010), 'Electrostatic actuator probe with curved electrodes for time-of-flight scanning force microscopy', *Review of Scientific Instruments*, vol. **81**, no. 8, pp. 083702 (6pp).
131. F. Hu, J. Yao, C. Qiu and H. Ren (June 2010), 'A MEMS micromirror driven by electrostatic force', *Journal of Electrostatics*, vol. **68**, no. 3, pp. 237–42.
132. L. Li, J. G. Brown and D. Uttamchandani (November 2002), 'Study of scratch drive actuator force characteristics', *Journal of Micromechanics and Microengineering*, vol. **12**, no. 6, pp. 736–41.
133. P. Honarmandi, J. W. Zu and K. Behdinan (May 2010), 'Analytical study and design characteristics of scratch drive actuators', *Sensors and Actuators A: Physical*, vol. **160**, no. 1–2, pp. 116–24.
134. H. Guckel, J. Klein, T. Christenson, K. Skrobis, M. Laudon and E. G. Lovell (1992), 'Thermo-magnetic metal flexure actuators', in *5th Technical Digest, IEEE Solid-State Sensor and Actuator Workshop, 1992.*, Hilton Head Island, SC, USA, pp. 73–5.

135. T. Moulton and G. K. Ananthasuresh (May 2001), 'Micromechanical devices with embedded electro-thermal-compliant actuation', *Sensors and Actuators A: Physical*, vol. **90**, no. 1–2, pp. 38–48.
136. C. S. Pan and W. Hsu (March 1997), 'An electro-thermally and laterally driven polysilicon microactuator', *Journal of Micromechanics and Microengineering*, vol. **7**, no. 1, pp. 7–13.
137. D. Yan, A. Khajepour and R. Mansour (July 2004), 'Design and modeling of a MEMS bidirectional vertical thermal actuator', *Journal of Micromechanics and Microengineering*, vol. **14**, no. 7, pp. 841–50.
138. S. M. Karbasi, M. Shamshirsaz, M. Naraghi and M. Maroufi (July 2010), 'Optimal design analysis of electrothermal microactuators', *Microsystem Technologies*, vol. **16**, no. 7, pp. 1065–71.
139. K. Mølhave and O. Hansen (June 2005), 'Electro-thermally actuated microgrippers with integrated force-feedback', *Journal of Micromechanics and Microengineering*, vol. **15**, no. 6, pp. 1265–70.
140. Q. -A. Huang and N. K. S. Lee (March 1999), 'Analysis and design of polysilicon thermal flexure actuator', *Journal of Micromechanics and Microengineering*, vol. **9**, no. 1, pp. 64–70.
141. Y. Zhu, A. Corigliano and H. D. Espinosa (January 2006), 'A thermal actuator for nanoscale in situ microscopy testing: design and characterization', *Journal of Micromechanics and Microengineering*, vol. **16**, no. 2, pp. 242–53.
142. P. Ginet, C. Lucat, F. Ménil and J. -L. Battaglia (October 2007), 'Modelling and characterizing a screen-printed metallic electrothermal microactuator', *International Journal of Applied Ceramic Technology*, vol. **4**, no. 5, pp. 423–7.
143. V. A. Henneken, M. Tichem and P. M. Sarro (March 2008), 'Improved thermal U-beam actuators for micro-assembly', *Sensors and Actuators A: Physical*, vol. **142**, no. 1, pp. 298–305.
144. L. -N. Tsai, G. -R. Shen, Y. -T. Cheng and W. Hsu (February 2006), 'Performance improvement of an electrothermal microactuator fabricated using Ni-diamond nanocomposite', *Journal of Microelectromechanical Systems*, vol. **15**, no. 1, pp. 149–58.
145. S. -C. Chen and M. L. Culpepper (October 2006), 'Design of contoured microscale thermomechanical actuators', *Journal of Microelectromechanical Systems*, vol. **15**, no. 5, pp. 1226–34.
146. J. Varona, M. Tecpoyotl-Torres and A. A. Hamoui (June 2009), 'Design of MEMS vertical–horizontal chevron thermal actuators', *Sensors and Actuators A: Physical*, vol. **153**, no. 1, pp. 127–30.
147. N. -T. Nguyen, S. -S. Ho and C. L. -N. Low (July 2004), 'A polymeric microgripper with integrated thermal actuators', *Journal of Micromechanics and Microengineering*, vol. **14**, no. 7, pp. 969–74.
148. Gih Keong Lau, J. F. L. Goosen, F. van Keulen, Trinh Chu Duc and P. M. Sarro (August 2008), 'Polymeric Thermal microactuator with embedded silicon skeleton: Part I – design and analysis', *Journal of Microelectromechanical Systems*, vol. **17**, no. 4, pp. 809–22.
149. Trinh Chu Duc, Gih-Keong Lau and P. M. Sarro (August 2008), 'Polymeric thermal microactuator with embedded silicon skeleton: Part II – fabrication, characterization, and application for 2-DOF microgripper', *Journal of Microelectromechanical Systems*, vol. **17**, no. 4, pp. 823–31.

150. N. Chronis and L. P. Lee (August 2005), ‘Electrothermally activated SU-8 microgripper for single cell manipulation in solution’, *Journal of Microelectromechanical Systems*, vol. **14**, no. 4, pp. 857–63.
151. T. Chu Duc, G. K. Lau and P. M. Sarro (September 2007), ‘Polymer constraint effect for electrothermal bimorph microactuators’, *Applied Physics Letters*, vol. **91**, no. 10, pp. 101902 (3pp).
152. Gih-Keong Lau, J. F. L. Goosen and F. van Keulen (August 2010), ‘Optimum design of polymeric thermal microactuator with embedded silicon skeleton’, *Journal of Microelectromechanical Systems*, vol. **19**, no. 4, pp. 992–1001.
153. W. -C. Chen, C. -C. Chu, J. Hsieh and W. Fang (January 2003), ‘A reliable single-layer out-of-plane micromachined thermal actuator’, *Sensors and Actuators A: Physical*, vol. **103**, no. 1–2, pp. 48–58.
154. M. Boutchich, T. J. Mamtora, G. J. McShane, I. Haneef, D. F. Moore and J. A. Williams (2006), ‘Force measurements on U-shaped electrothermal microactuators: applications to packaging’, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. **222**, no. 1, pp. 87–96.
155. A. A. Geisberger and N. Sarkar (2006), ‘Techniques in MEMS microthermal actuators and their applications’, in C. T. Leondes (Ed.), *MEMS/NEMS Handbook Techniques and Applications Volume 4 Sensors and Actuators*, ch. 8, Springer, pp. 239–54.
156. L. Chen, C. Liu, K. Liu, C. Meng, C. Hu, J. Wang and S. Fan (March 2011), ‘High-performance, low-voltage and easy-operable bending actuator based on aligned carbon nanotube/polymer composites’, *American Chemical Society Nano*, vol. **5**, no. 3, pp. 1588–93.
157. B. Rogers, L. Manning, T. Sulchek and J. D. Adams (August 2004), ‘Improving tapping mode atomic force microscopy with piezoelectric cantilevers’, *Ultramicroscopy*, vol. **100**, no. 3–4, pp. 267–76.
158. R. Maeda, J. J. Tsau, S. H. Lee and M. Ichiki (January 2004), ‘Piezoelectric microactuator devices’, *Journal of Electroceramics*, vol. **12**, no. 1–2, pp. 89–100.
159. Z. Qiu, J. S. Pulskamp, X. Lin, C. -H. Rhee, T. Wang, R. G. Polcawich and K. Oldham (June 2010), ‘Large displacement vertical translational actuator based on piezoelectric thin films’, *Journal of Micromechanics and Microengineering*, vol. **20**, no. 7, pp. 075016 (10pp).
160. Y. Zhu, W. Liu, K. Jia, W. Liao and H. Xie (2011), ‘A piezoelectric unimorph actuator based tip-tilt-piston micromirror with high fill factor and small tilt and lateral shift’, *Sensors and Actuators A: Physical*, vol. **167**, no. 2, pp. 495–501, doi:10.1016/j.sna.2011.03.018.
161. M. Rakotondrabe and I.A. Ivan (December 2010), ‘Development and dynamic modeling of a new hybrid thermopiezoelectric microactuator’, *IEEE Transactions on Robotics*, vol. **26**, no. 6, pp. 1077–85.
162. J. C. Doll, B. C. Petzold, B. Ninan, R. Mullapudi and B. L. Pruitt (February 2010), ‘Aluminum nitride on titanium for CMOS compatible piezoelectric transducers’, *Journal of Micromechanics and Microengineering*, vol. **20**, no. 2, pp. 025008 (8pp).
163. A. Andrei, K. Krupa, M. Jozwik, P. Delobelle, L. Hirsinger, C. Gorecki, L. Nieradko and C. Meunier (February 2008), ‘AlN as an actuation material for MEMS applications: the case of AlN driven multilayered cantilevers’, *Sensors and Actuators A: Physical*, vol. **141**, no. 2, pp. 565–76.

164. S. Li and S. Chen (April 2003), 'Analytical analysis of a circular PZT actuator for valveless micropumps', *Sensors and Actuators A: Physical*, vol. **104**, no. 2, pp. 151–61.
165. A. Doll, M. Wischke, H. -J. Schrag, A. Geipel, F. Goldschmidtboeing and P. Woias (May–August 2007), 'Characterization of active silicon microvalves with piezoelectric membrane actuators', *Microelectronic Engineering*, vol. **84**, no. 5–8, pp. 1202–6.
166. T. Mashimo and S. Toyama (August 2010), 'Rotary-linear piezoelectric micro-actuator with a cubic stator of side length 3.5 mm', *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. **57**, no. 8, pp. 1825–30.
167. C. -H. Yun, B. Watson, J. Friend and L. Yeo (August 2010), 'A piezoelectric ultrasonic linear micromotor using a slotted stator', *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. **57**, no. 8, pp. 1868–74.
168. Eunki Hong, S. Trolier-McKinstry, R. L. Smith, S. V. Krishnaswamy and C. B. Freidhoff (August 2006), 'Design of MEMS PZT circular diaphragm actuators to generate large deflections', *Journal of Microelectromechanical Systems*, vol. **15**, no. 4, pp. 832–9.
169. C. Bolzmacher, K. Bauer, U. Schmid, M. Hafez and H. Seidel (January 2010), 'Displacement amplification of piezoelectric microactuators with a micro-machined leverage unit', *Sensors and Actuators A: Physical*, vol. **157**, no. 1, pp. 61–7.
170. P. R. Ouyang, R. C. Tjiptoprodjo, W. J. Zhang and G. S. Yang (August 2008), 'Micro-motion devices technology: the state of arts review', *International Journal of Advanced Manufacturing Technology*, vol. **38**, no. 5–6, pp. 463–78.
171. Y. -C. Tsai, S. H. Lei and H. Sudin (January 2005), 'Design and analysis of planar compliant microgripper based on kinematic approach', *Journal of Micromechanics and Microengineering*, vol. **15**, no. 1, pp. 143–56.
172. O. Sardan, D. H. Petersen, K. Mølhave, O. Sigmund and P. Bøggild (May–June 2008), 'Topology optimized electrothermal polysilicon microgrippers', *Microelectronic Engineering*, vol. **85**, no. 5–6, pp. 1096–9.

12

Dynamic behavior of smart MEMS in industrial applications

M. PUSTAN, C. BIRLEANU and C. DUDESCU,
Technical University of Cluj-Napoca, Romania and
J-C. GOLINVAL, University of Liège, Belgium

DOI: 10.1533/9780857099297.2.349

Abstract: Many of the microelectromechanical systems (MEMS) industrial applications require vibrating components that operate to a high quality factor and small energy dissipation during oscillations. To improve the reliability design of MEMS resonators, the effect of operating conditions on the dynamical response of vibrating components has to be accurately determined. As a function of the operating conditions, the dynamical response and the loss of energy in vibrating MEMS components are influenced by the damping of the surrounding medium and depend on the intrinsic effects of mechanical structure. In order differentiate between the extrinsic damping and the intrinsic effects, experiments have to be performed both in ambient conditions and in a vacuum. In this chapter, some analytical models accompanied by experimental tests are presented to estimate the dynamical response and the loss of energy on samples fabricated from polysilicon with different geometrical configurations.

Key words: MEMS resonators, dynamic response, electromechanical coupling, quality factor, energy dissipation.

12.1 Introduction

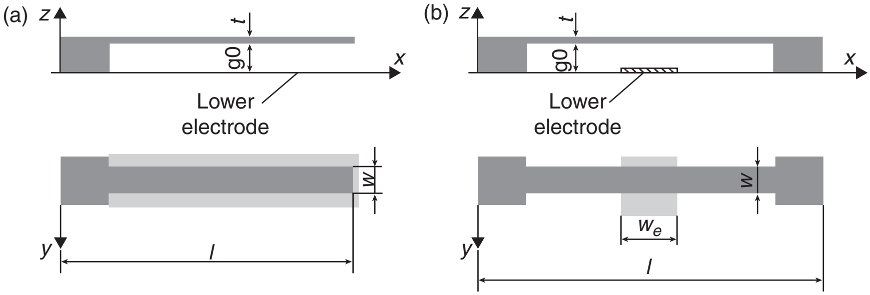
Modeling and experimental tests of resonators from microelectromechanical systems (MEMS) are essential for design optimization and device reliability. One of the most important applications of MEMS resonators is mass-detection for chemical and biological applications, radio frequency applications (RF MEMS), in the automobile industry and for aircraft conditions monitoring or satellite communications. The reliability and the lifetime are crucial parameters in different MEMS applications and these are strongly dependent on the material properties. Moreover, MEMS resonators need to be designed to perform their expected function with complete accuracy. The accuracy of a response is influenced by a resistor's material properties. As a function of applications, MEMS resonators operate under various conditions including different temperatures, humidity or pressure of the surrounding medium. The mechanical behavior of resonators strongly depends on the operating conditions. In order to design a reliable MEMS resonator, experimental investigations are needed to evaluate the accuracy of their mechanical response under different loadings and operating conditions.

Two of the major causes of failure in MEMS resonators which operate under high-cycle loading are fatigue and the loss of energy during vibration, based on thermal effects. The high-cycle fatigue life is greatly influenced by microstructural variables such as the grain size, the volume fraction of the secondary phase and the amount of solute atoms or precipitates. The information related to fatigue failure comes from Stress-Numbers of cycles (S-N) curves, which is the traditional way of representing fatigue data. For cyclic motions of a structural material, significant heat generation occurs and energy dissipation is produced due to an energy loss mechanism internal to the material. The temperature gradient generates heat currents which cause an increase in entropy in the resonator and lead to energy dissipation. It would be highly desirable to design a MEMS resonator with as little loss of energy as possible. Experimentally, the loss of energy in MEMS resonators is evaluated considering the frequency response curves and measuring the bandwidth of oscillations under an excitation signal. In this chapter, experimental investigations are performed in order to estimate the loss of energy in MEMS resonators. Most of the MEMS vibration sensors that were used have a polysilicon microcantilever or microbridge as the sensing element. Even though these components are simple geometrical structures, their dynamical behavior needs to be more accurately investigated.

12.2 Resonant frequency response of smart MEMS vibrating structures

Mechanical resonators such as microcantilevers and microbridges are very often used as flexible mechanical components in MEMS. There are many applications that require ambient operating conditions and others that require working in a vacuum. As a consequence, during experimental investigations, the samples are successively tested in air and in a vacuum, and the effect of the surrounding medium on the amplitude and velocity of oscillations is estimated accordingly. The dynamic response of samples is changed as a function of the operating conditions.

In this chapter, vibrating MEMS resonators such as microcantilevers (Fig. 12.1a) and microbridges (Fig. 12.1b) are dynamically investigated and their frequency responses under a harmonic loading is determined under different testing conditions.



12.1 Schematic of a microcantilever and a microbridge under electrostatic actuation. (a) Microcantilever, (b) microbridge.

Mechanical stiffness is a fundamental criterion of elastically deformable mechanical flexible microcomponents. The bending stiffness of a microcantilever and a microbridge can be computed using the following well-known relations (Lobontiu and Garcia, 2004; Pustan, 2007):

- for a microcantilever

$$k_{\text{cantilever}} = \frac{1}{4}\beta \cdot E \tag{12.1}$$

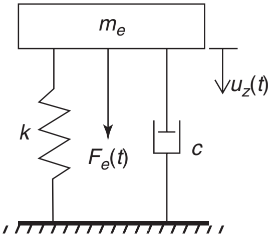
- for a microbridge

$$k_{\text{bridge}} = 16\beta \cdot E \tag{12.2}$$

where E is the Young's modulus of material and $\beta = wt^3/l^3$ is the resonator geometrical parameter given by w = width, t = thickness and l = length of samples.

When a DC voltage (V_{DC}) is applied between the lower electrode and the vibrating MEMS structure, an electrostatic force is set up and the cantilever bends downwards and comes to rest in a new position. To drive the resonator at resonance, an AC harmonic load of amplitude V_{AC} vibrates the cantilever in the new deflected position.

A single degree of freedom model, as presented in Fig. 12.2, can be used to simulate the dynamic response of a resonator due to the V_{DC} and V_{AC} electric loadings. In this model, the proof mass of the cantilever is modeled as a lumped mass m_e, and its stiffness is considered as a spring constant k. This part forms the vibrating side of a variable capacitor. The bottom electrode is



12.2 Forced vibration model with fixed support used in dynamic investigations.

fixed and considered as the second part of the MEMS structure. If a voltage composed of DC and AC terms as:

$$V = V_{\mathrm{DC}} + V_{\mathrm{AC}} \cos(\omega t) \qquad [12.3]$$

is applied between the resonator electrodes, the electrostatic force applied on the structure has a DC component as well as a harmonic component with the frequency ω such that:

$$F_e(t) = \frac{\varepsilon A V^2}{2\left[g_0 - u_z(t)\right]^2} \qquad [12.4]$$

where ε is the permittivity of the free space, $A = w_e \times w$ is the effective area of the capacitor, g_0 is the initial gap between the flexible plate and the substrate and $u_z(t)$ is the displacement of the mobile plate under the electrostatic force $F_\mathrm{e}(t)$.

The expression [12.4] evidences two aspects: the electromechanical coupling between the instantaneous value of beam gap ($g_0 - u_z$) and the applied voltage, then the nonlinear dependence between the mechanical displacement u_z and the voltage.

Pull-in voltage, at which the elastic stiffness does not balance the electric actuation and the beam tends to collapse, can be evaluated by establishing the maximum gap allowing the static equilibrium. The spring force and the electrostatic actuation have opposing directions. The instability threshold is found by imposing the two conditions of null total force and the null first derivative with respect to the displacement:

$$k u_z - \frac{\varepsilon A V^2}{2\left(g_0 - u_z\right)^2} = 0 \qquad [12.5]$$

$$k-\frac{\varepsilon A V^2}{2\left(g_0-u_z\right)^3}=0 \qquad [12.6]$$

The unknown displacement and voltage are:

$$u_{\text{pull-in}}=\frac{g_0}{3} \qquad [12.7]$$

$$V_{\text{pull-in}}=\sqrt{\frac{8}{27}\frac{g_0{}^3 k}{\varepsilon A}} \qquad [12.8]$$

where $u_{\text{pull-in}}$ and $V_{\text{pull-in}}$ are the maximum displacement and voltage at which is possible to have a stable equilibrium configuration, k is the beam stiffness described by Equation [12.1] for a microcantilever and Equation [12.2] for a microbridge.

Dynamic analysis of electrostatically actuated microcomponents is performed by linearizing the electrostatic actuation around an equilibrium position. The equivalent stiffness of a MEMS resonator can be computed as:

- for a microcantilever

$$k_{\text{eff}}=\frac{3EI_y}{l^3}-\frac{\varepsilon A V^2}{\left(g_0-u_z\right)^3} \qquad [12.9]$$

- for a microbridge

$$k_{\text{eff}}=\frac{192EI_y}{l^3}-\frac{\varepsilon A V^2}{\left(g_0-u_z\right)^3} \qquad [12.10]$$

Based on these equations, the resonant frequency of an electrostatically actuated microcantilever and microbridge can be computed as:

$$\omega_0=\frac{1}{2\pi}\sqrt{\frac{k_{\text{eff}}}{m_e}} \qquad [12.11]$$

where m_e is the equivalent mass of system.

Using the assumption that the kinetic energy of the distributed-parameter system is equal to the kinetic energy of the equivalent lumped-parameter mass, the equivalent mass can be determined (Lobontiu, 2007). The equivalent mass of a microcantilever is 0.235 m, and of a microbridge is 0.406 m (m is the effective mass of beam).

The dynamic response of MEMS resonators presented in Fig. 12.1 subjected to a harmonic electrostatic force $F_e(t)$ with the driving frequency ω given by an AC voltage is governed by the equation of motion:

$$m\cdot\ddot{u}_z(t)+c\cdot\dot{u}_z(t)+k\cdot u_z(t)=F_e(t) \qquad [12.12]$$

where c is the damping factor.

The response of the system under DC and AC voltages is given by the equation:

$$u_z(t)=\frac{u_z}{\sqrt{\left(1-\left(\frac{\omega}{\omega_0}\right)^2\right)^2+\left(2\xi\frac{\omega}{\omega_0}\right)^2}} \qquad [12.13]$$

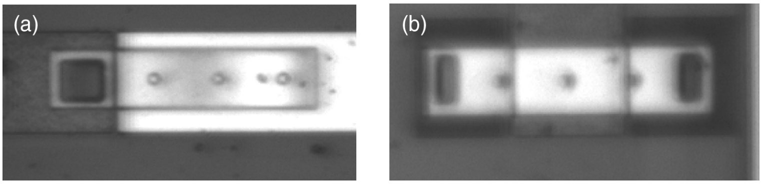
where ξ is the damping ratio and ω_0 is the resonant frequency of beams given by Equation [12.11].

Usually, the response is plotted as a normalized quantity $u_z(t)/u_z$. When the driving frequency equals the resonant frequency $\omega=\omega_0$, the amplitude ratio reaches a maximum value. At resonance, the amplitude ratio becomes:

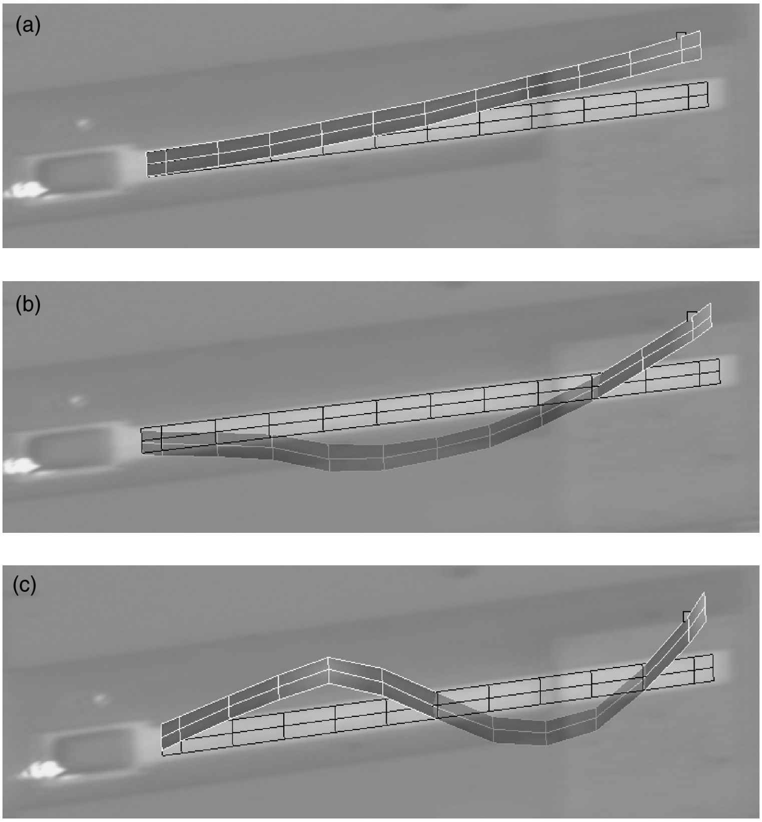
$$\frac{u_z(t)}{u_z}=\frac{1}{2\xi} \qquad [12.14]$$

The experimental investigations of the vibrating MEMS structures are performed using a vibrometer analyzer and a white noise signal. The aim of experimental investigations is to determine the frequency response of a microcantilever and a microbridge, and the effect of the operating conditions on the velocity and amplitude of oscillations.

The geometrical dimensions of the microresonators from Fig. 12.3 are the following: total length l of beams is 150 µm; width w is 30 µm and thickness t is 1.9 µm; the gap between the flexible plates and the substrate g_0 is 2 µm; the holes have a diameter of 3 µm; the width w_e of the lower electrode of microbridge is 50 µm. The microcantilever is fabricated with the full lower electrode under the flexible plate. During tests, a DC offset signal

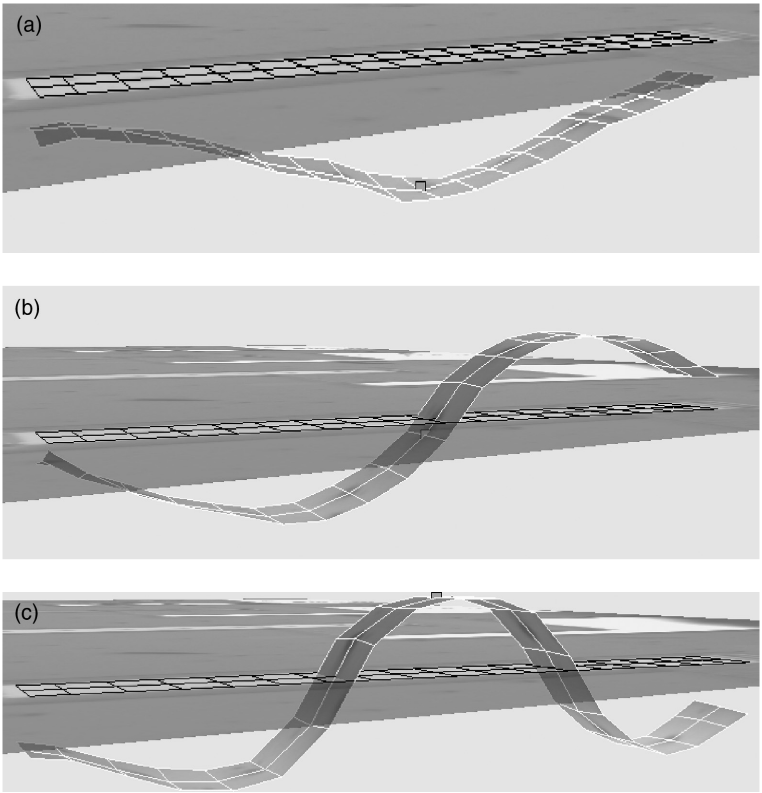


12.3 Microresonators used in experimental investigations. (a) Microcantilever and (b) microbridge.



12.4 The bending modes of oscillations of an electrostatically actuated MEMS cantilever. (a) Bending mode 1, (b) bending mode 2 and (c) bending mode 3.

of 5 V and peak amplitude of 5 V of the driving signal are applied to bend and oscillate the samples. The frequency response, and the amplitude and velocity of oscillations are measured under continuous actuation of the microresonator.



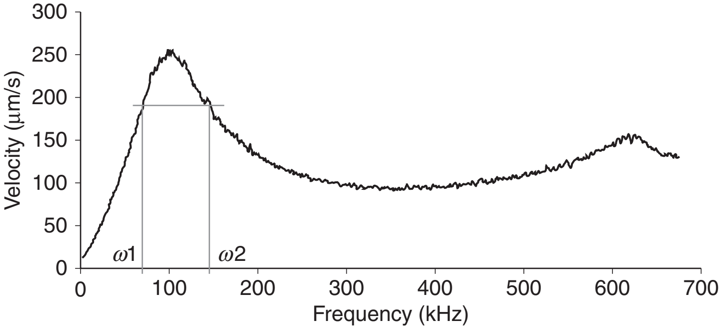
12.5 The bending modes of oscillations of an electrostatically actuated MEMS microbridge. (a) Bending mode 1, (b) bending mode 2 and (c) bending mode 3.

The frequency responses of the samples investigated can be monitored for different oscillation modes using a vibrometer analyzer. As presented in Figs 12.4 and 12.5, three bending modes of oscillation were monitored under an excitation signal.

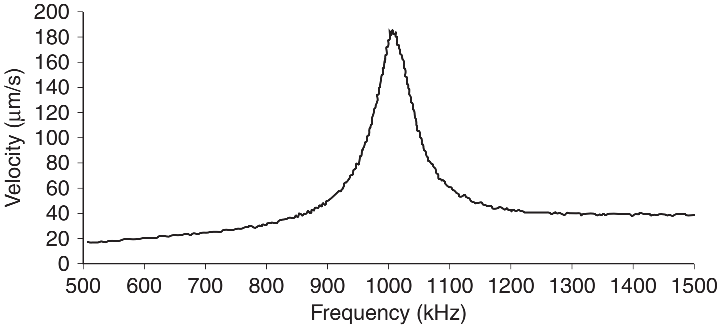
The tests are performed under ambient conditions and in a vacuum in order to estimate the damping effect on the velocity and amplitude of oscillations.

In order to analyze the dynamic response of the MEMS resonators under investigation, only the first bending mode is monitored and analyzed. The frequency response curves of a microcantilever tested in air are presented in Fig. 12.6 and those of microbridge in Fig. 12.7. Figures 12.8 and 12.9 present the frequency response of the same resonators tested in a vacuum.

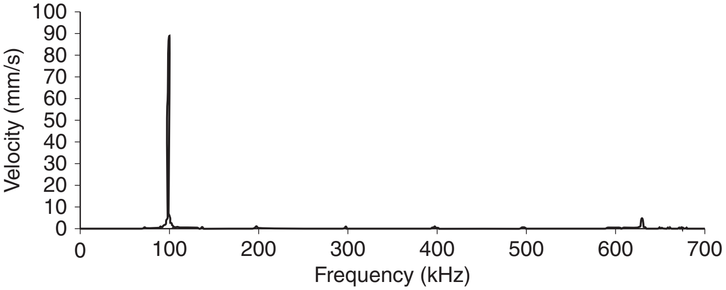
The dynamic experimental characteristics of the investigated microbridge and microcantilever are presented in Table 12.1.



12.6 Frequency response of an electrostatically actuated MEMS microcantilever tested in ambient conditions.

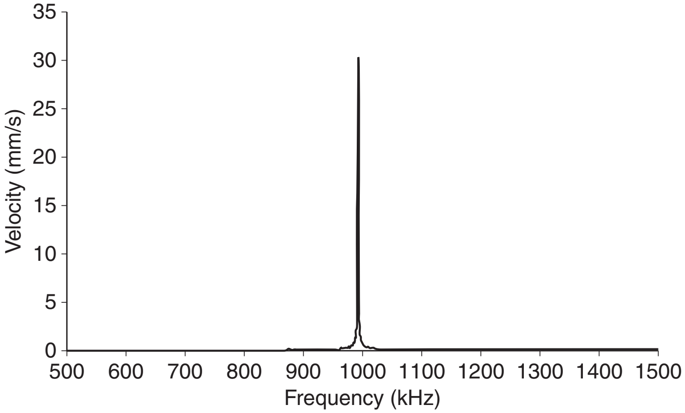


12.7 Frequency response of an electrostatically actuated MEMS microbridge tested in ambient conditions.



12.8 Frequency response of an electrostatically actuated MEMS microcantilever tested in a vacuum. Vacuum at 8 × 10^{-6} mbar.

The experimental results of the MEMS resonators concur with the analytical models presented above. As can be observed in Table 12.1, there are small differences between frequency responses of beams tested in different operating conditions. These differences depend on the damping of the



12.9 Frequency response of oscillations of an electrostatically actuated MEMS microbridge tested in a vacuum. Vacuum at 8 × 10^{-6} mbar.

Table 12.1 Dynamic experimental characteristics function of testing condition

Resonator type	Resonant frequency (kHz)		Velocity (mm/s)		Amplitude (nm)	
	Air	Vacuum	Air	Vacuum	Air	Vacuum
Cantilever	100	99.37	0.27	88	0.47	140
Bridge	1003.37	992.81	0.19	31	0.03	4.96

surrounding medium, that causes a shift in the frequency response of the beam. Significant differences were observed in the velocity and amplitude of oscillations. The amplitude and velocity of oscillations have small values if the microresonators are tested in ambient conditions based on the damping of the surrounding medium. The air damping changes not only the dynamical characteristics such as resonant frequency, amplitude and velocity of oscillations, but also the quality factor and the loss coefficient of energy, as presented in next section.

12.3 Quality factor and the loss coefficient of smart MEMS vibrating structures

The energy dissipated during one cycle of oscillation can be evaluated based on the quality factor Q. The quality factor is an important qualifier of mechanical microresonators and allows estimation of the loss coefficient of oscillations $Q^{-1} = 1/Q$. In terms of energy, it is expressed as the total energy stored in the system divided by the energy dissipated per cycle. At resonance, the quality factor is expressed as (Lobontiu, 2007):

$$Q_r = \frac{1}{2\xi} \quad [12.15]$$

and the normalized response given by Equation [12.14] is equal to Q_r.

The quality factor is also known as 'sharpness at resonance', which is defined as the ratio:

$$Q_r = \frac{\omega}{\omega_2 - \omega_1} \quad [12.16]$$

where $\omega_2 - \omega_1$ is the frequency bandwidth corresponding to 0.707 $u_z(t)_{\max}$ on the amplitude *versus* resonant frequency curves (as shown in Fig.12.6).

The total loss coefficient occurring in a microresonator can be separated into two components as:

$$Q^{-1}_{\text{total}} = Q_e^{-1} + Q_i^{-1} \quad [12.17]$$

where *e* denotes the extrinsic losses and *i* the intrinsic losses.

Some of the extrinsic mechanisms are affected by changes of environment. The air damping can be minimized under ultrahigh vacuum conditions. Intrinsic losses in the resonator material are an important mechanism in accounting for energy dissipation.

A small internal loss is produced by the energy dissipation anchors that attach the resonator to substrate. The clamping losses can be determined by analyzing the vibration energy which is transmitted from resonator to substrate, and for one anchor it can be computed as (Lobontiu, 2007):

$$Q^{-1}_{\text{anchor}} = 2.17\left(\frac{0.5l}{t}\right)^3 \quad [12.18]$$

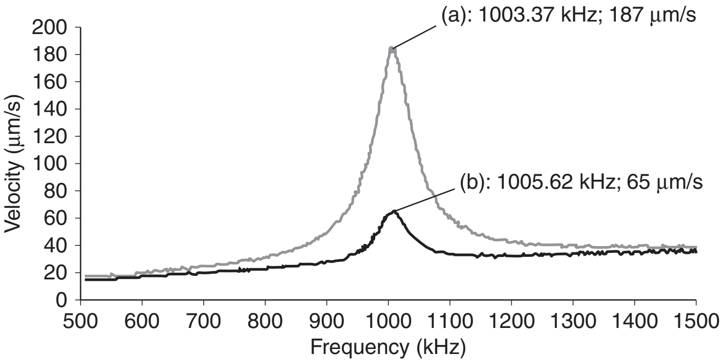
where *l* is the length and *t* is the thickness of resonators.

For cyclic motions of a structural material, significant heat generation become apparent and energy dissipation occurs due to an energy loss mechanism internal to the material (Lobontiu, 2007). The variation of strain in a microresonator is accompanied by a variation of temperature, which causes an irreversible flow of heat. The temperature gradient generates heat currents which cause an increase in the entropy of the beam and lead to energy dissipation. This process of energy dissipation is known as 'thermoelastic damping'. Thermoelastic damping depends on material properties such as the specific heat, coefficient of thermal expansion, thermal conductivity, mass density and elastic modulus, as well as the temperature and geometry. Thermoelastic damping is recognized as an important loss mechanism at room temperature in micro-scale beam resonators. The mechanism of thermoelastic damping was first studied by Zener (1937) and later developed in Lifshtz and Roukes (2000) and Yi (2008). He indicates that the phenomenon is induced by irreversible heat dissipation during the coupling of heat transfer and the strain rate in an oscillating system. The Zener model used the classical thermoelastic theory assuming an infinite speed of heat transmission. In a more complex model (Sun *et al.*, 2006) based on generalized thermoelastic theory with one relaxation time, the bending moment on the beam during oscillations is separated into two parts: the first is the well-known moment which arises from the bending of the beam when the temperature gradient across the beam is zero; the second moment is the bending moment which arises from the variation of temperature across the upper and lower surfaces of the beam known as the 'thermal moment'. Analytical results (Sun *et al.*, 2006) shown that thermoelastic coupling influences the amplitude, velocity and resonant frequency of a beam based on its thermal moment. Over time, the deflection and thermal moment attenuate. The energy dissipation in a microresonator is given by means of the thermal moment variation followed by the attenuation of the amplitude (Sun *et al.*, 2006). The theoretical results were validated by experiments (Pustan *et al.*, 2012).

The total loss coefficient is experimentally determined when the sample oscillates in ambient conditions. The sample response in a vacuum determines the intrinsic losses. For the microresonators with the geometrical dimensions presented above (p. 354), the experimental tests are performed both in a vacuum and in ambient conditions. Using the frequency bandwidth ($\omega_2 - \omega_1$) corresponding to $0.707u_z(t)_{\max}$ on the frequency response experimental curves presented in Figs 12.6–12.9, the quality factor Q and the loss coefficient Q^{-1} are determined and presented in Table 12.2.

Table 12.2 Quality factors Q and loss coefficient Q^{-1} of investigated microresonators

Resonator type	Quality factor Q		Loss coefficient Q^{-1}	
	Air	Vacuum	Air	Vacuum
Cantilever	1.33	310.5	75×10^{-2}	32.2×10^{-4}
Bridge	26.78	2239.69	3.73×10^{-2}	4.46×10^{-4}



12.10 Frequency responses of a microbridge resonator in ambient conditions. Depicted in a frequency domain from 500 to 1500 kHz: (a) the initial response and (b) the beam response after 4 h.

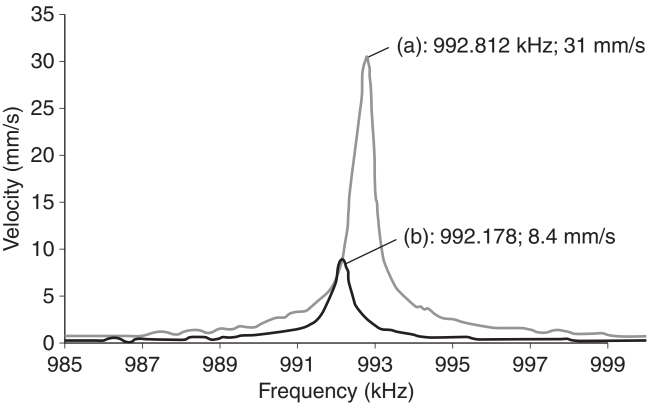
Using Equation [12.15], the damping ratio of tested samples in ambient conditions can be estimated. A damping ratio of 0.375 is determined for the microcantilever and 0.018 for the microbridge. The damping ratio ξ is any positive real number. For the value of the damping ratio $0 \leq \xi < 1$ as in the experiments, the system has an oscillatory response.

Experimental tests are conducted in order to estimate the thermomechanical coupling effect on the vibrating structures as function of operating time. The following presents the case of a microbridge resonator.

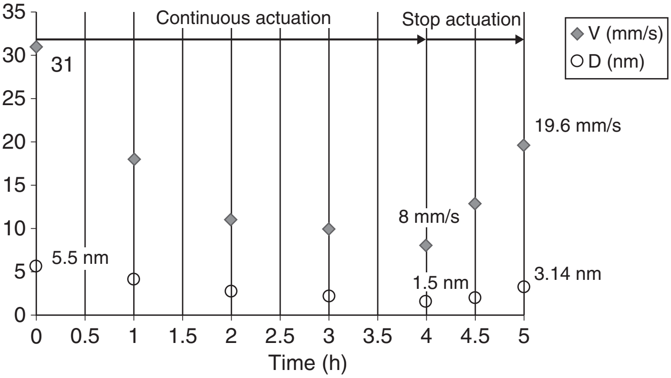
The velocity of oscillations in ambient conditions decreases from 187 to 65 μm/s after 4 h (Fig. 12.10); in a vacuum, the velocity is attenuated from 31 to 8.4 mm/s (Fig. 12.11). The same decreases in the microbridge resonator displacements as a function of the operating time were also observed.

Figure 12.12 shows the attenuation of velocity and displacement as a function of the oscillating time. The microresonator oscillated continuously for 4 h and the changes in its dynamic response were observed at hourly intervals. After 4 h the excitation of the sample was stopped. The next test was commenced after 30 min, when the increase in the velocity and displacement of oscillations was observed. After 1 h with no actuation of the beam, the thermal effect decreases and the beam response is improved. The velocity of oscillations increases from 8 to 19.6 mm/s and displacement from 1.5 to 3.14 nm (Fig. 12.12).

The tests were repeated three times (in different days) and the same attenuation of velocity and amplitude was observed. The average attenuation of velocity and displacement is about 65%. The attenuation in velocity and amplitude of oscillations are based on the thermoelastic coupling and change of the thermal moment as reported by Sun *et al.* (2006). The same analytical study revealed that the computed thermal moment is attenuated



12.11 Frequency responses of a microbridge resonator in vacuum. Depicted in a frequency range from 985 to 999 kHz: (a) the initial response and (b) the beam response after 4 h; vacuum at 8 × 10^{-6} mbar.



12.12 Experimental variation of velocity *V* (mm/s) and displacement *D* (nm) in vacuum as a function of operating time.

significantly after a longer time and the deflection amplitude (peak value) decreases by about 50% after an operating time range because the effect of thermoelastic damping is enhanced. Also, over time, the prestressed position given by DC current is changed based on the thermal relaxation of the material; it has an influence on the forces balance equation and on the peak amplitude of oscillation described by Equation [12.13].

The thermoelastic effect changes the resonant frequency as presented in Figs 12.10 and 12.11. The air damping effect can increase the frequency response due to the change of the medium compressibility factor. The air escapes from the gap formed between the movable and fixed components, its compressibility generating the spring behavior (Yi, 2008). The compressibility factor changes with temperature. During testing, the heat propagation from a vibrating sample changes the temperature of the surrounding medium, decreasing the compressibility factor of the medium. As a consequence, extrinsic damping decreases and changes the resonant frequency of the beam (Ahmadian *et al.*, 2009).

The total loss coefficient is experimentally determined when the sample oscillates in ambient conditions. The sample response in vacuum determines the intrinsic losses. Table 12.3 shows the quality factors of a microbridge resonator at the beginning of its operating time (Q_{0h}) and after 4 h (Q_{4h}). The changes in the quality factor as a function of operating time can be observed.

Table 12.4 shows the changes of the loss coefficient of energy as a function of operating time. The experiments were repeated three times (on different days) and the same change (13% increasing) of thermoelastic losses was obtained. The increases in the total loss coefficient $Q^{-1}{}_{tot}$ were different for each day (26%, 19%, 22%) because of the changes in environmental conditions (the ambient conditions were not controlled during testing). The environmental conditions have a considerable influence on the extrinsic loss coefficient when the sample is tested in ambient conditions.

The strain energy method is used in ANSYS/Multiphysics to compute the loss coefficient. For the microbridge resonator, a loss coefficient of 5.1×10^{-4} was determined (Pustan *et al.*, 2012), a value that was close to the experimental measurement.

Table 12.3 Dependence of quality factors on operating time

Quality factor	Testing conditions	
	Air	Vacuum
Q_{0h}	26.78	2239.69
Q_{4h}	19.86	1943.66

Table 12.4 Dependence of loss coefficients on operating time

Loss coefficient	Initial	After 4 h
Q^{-1}_{total}	3.7×10^{-2}	5×10^{-2}
Q^{-1}_{i}	4.46×10^{-4}	5.144×10^{-4}
Q^{-1}_{e}	3.6×10^{-2}	4.9×10^{-2}

12.4 Industrial applications: resonant accelerometers

Resonant accelerometers have many applications in automobiles, inertial navigation systems, avionics and satellites. They detect external acceleration by measuring the frequency variation of a resonating part. The basic structures for resonant vibration are microcantilevers and microbridges, as shown in Fig. 12.3. Each structure has several different resonant modes, where each mode has its own displacement pattern, resonant frequency and *Q*-factor. The advantage of using frequency shift as the sensing parameter is its high-level signal and lower sensitivity to parasitic influences. The resonant accelerometer is based on a differential capacitive sensing structure, as previously described in this section. The resonant sensor is an element vibrating at resonance. The shift of frequency is a function of the parameter to be measured, acting on the resonator by changing properties such as the shape of the sensor, inducing stress or adding mass (Kempe, 2011). These are directly related to the frequency of the sensor. The advantage of using frequency as the sensing parameter is its high-level signal and lower sensitivity to parasitic influences.

The mechanical resonator sensors must be set in one of the vibration modes (Figs 12.4 or 12.5), and the vibration detected. One of the main types of excitation technique is electrostatic actuation and the capacitive detection method as presented in this chapter. This is a vibration excitation technique for a resonator oscillated in both in a vacuum and air. For the resonators operated in air, it is critical to design the movable electrode to ensure free movement of the air. Air damping is reduced if the electrodes have holes through which the air is free to move.

The dynamical response and the loss of energy in vibrating MEMS components is influenced by the damping of the surrounding medium and depends on the intrinsic effects of mechanical structure. The air damping changes not only the dynamical response as resonant frequency, amplitude and velocity of oscillations, but also decreases the quality factor and increases the loss coefficient of energy. For cyclic motions of a structural material, significant heat generation becomes apparent and energy dissipation occurs due to an energy loss mechanism within the material. The attenuation in velocity and amplitude of oscillations under continuous actuation are based on the thermoelastic coupling and change of the thermal moment as presented in this chapter.

12.5 References

Ahmadian, M.T., Borhan, H. and Esmailzadeh, E. (2009) 'Dynamic analysis of geometrically nonlinear and electrostatically actuated micro-beams', *J. Nonlin. Sci. Num. Sim.*, **14**, 1627–1645.

Kempe, V. (2011) *Inertial MEMS: Principles and Practice*, New York: Cambridge University Press.

Lifshtz, R. and Roukes, M.L. (2000) 'Thermoelastic damping in micro-and nanomechanical systems', *Phys. Rev. B.*, **61**, 5600–5609.

Lobontiu, N. (2007) *Dynamics of Microelectromechanical Systems*, New York: Springer Science.

Lobontiu, N. and Garcia, E. (2004) *Mechanics of Microelectromechanical Systems*, Ithaca, NY: Cornell University.

Pustan, M., Birleanu, C. and Dudescu, C. (2012) 'Simulation and experimental analysis of thermo-mechanical behavior of microresonators under dynamic loading'. Symposium on Design, Test, Integration and Packaging of MEMS/MOEMS (DTIP 2012), Cannes, France, 25–27 April 2012, 87–92.

Pustan, M. and Rymuza, Z. (2007) 'Mechanical properties of flexible microcomponents with movable load', *J. Micromech. Microeng.,* **17**, 1611–1617.

Sun, Y., Fang, D. and Soh, A.K. (2006) 'Thermoelastic damping in micro-beam resonators', *Int. J. Sol. Str.*, **43**, 3213–3229.

Yi, Y.B. (2008) 'Geometrical effects on thermoelastic damping in MEMS resonators', *J. Sound. Vib.*, **309**, 588–599.

Zener, C. (1937) 'Internal friction in solids. Theory of internal friction in reeds', *Phys. Rev.*, **32**, 230–235.

13

MEMS integrating motion and displacement sensors

G. LANGFELDER and A. TOCCHIO,
Politecnico di Milano, Italy

DOI: 10.1533/9780857099297.2.366

Abstract: The purpose of this chapter is to describe the working principle of the micro electro mechanical systems (MEMS) devices used in motion sensors, and to consider their level of sophistication and future evolution. The integration of a multi-parameter sensor in a single unit can increase the smartness of the device and enables the implementation of new functionalities in existing electronic systems. MEMS accelerometers, gyroscopes and magnetometers are discussed in detail, as they represent the core devices for the development of a high-precision inertial measurement unit (IMU). At the end of the chapter, consideration is given to the evolution of these units integrating new MEMS devices for added functionalities, such as proximity measurements.

Key words: MEMS motion sensors, inertial measurement units, accelerometers, gyroscopes, magnetometers.

13.1 Introduction

A smart motion sensor is, in its simplest form, a single unit capable of providing information on linear and rotational displacements for all the three possible axes of motion. This kind of sensor can be presently made using micro electro mechanical systems (MEMS) technology for the transducing part, often combined with an integrated circuit that implements the control and readout electronics. The main purposes of this chapter are to present the basic working principle of the MEMS devices used to implement this unit, to describe typical and advanced driving and sensing methods, and to explore the evolution of their smartness in the near future.

The basic devices that make up a motion sensor unit, also referred to as an inertial measurement unit (IMU), are the accelerometer and the gyroscope; the accelerometer is used to measure linear motion, the gyroscope is used to measure rotational motion. These devices have been used for decades, but the extensive spread of their use since the turn of the millennium is due to the research and industrial progress obtained in MEMS technology. First, the possibility of implementing such devices in very small dimensions and with ultra-low power consumption was very attractive and led to the development of primitive MEMS devices for single-axis, single-parameter measurements (Boser *et al.*, 1996; Clark *et al.*, 1996). But it is only in the last few years (since 2009) that the smartness obtained from the combination of multi-axes and multi-parameter MEMS units has opened new fields of application for IMUs and contributed to the integration of new functionalities in existing electronic systems. This combination of sensors gives advantages in terms of both cost production for the MEMS supplier and solutions offered to the system integrator.

In particular, consumer electronics has profited enormously from this development and now tri-axial accelerometers and gyroscopes can be found in game controllers, sport equipment, notebooks and netbooks, portable media players, digital still cameras and mobile phones. Actually, it can be stated that most of the smartness of a smartphone is given by the sensors embedded in it. Among the other sectors, the automotive industry still holds almost the same market share as consumer electronics, with massive use in airbags, electronic stability control (ESC) and tire pressure monitoring systems (TPMS). Inertial sensors with higher performance requirements can be found also in the medical, industrial, aerospace and military markets (e.g. medical human motion analysis, land transportation systems, oil drilling and exploration, civil and military aviation and unmanned vehicles). On the whole, according to Robin *et al.* (2011), it is anticipated that the accelerometers and gyroscopes market will have grown from $2.3 billion in 2010 to $4 billion by 2016.

The state-of-the-art of IMUs made with MEMS sensors is only represented by the ‘6-axis’ (or ‘6-degrees-of-freedom (DOF’)) units, integrating 3-axial accelerometers and gyroscopes made in the same process (examples of which can be found in Analog Devices (2010) and STMicroelectronics (2011)). Further smartness and a greater number of features could be obtained if the IMU were also capable of measuring the absolute orientation of the object that incorporates it. With this added feature, heading, navigation, compass and dead-reckoning applications could be implemented because the loss of absolute orientation over time, which is typical of 6-DOF IMUs, could be corrected. Absolute orientation can be obtained using a magnetometer, or digital compass. At the state-of-the-art, 9-DOF systems integrate non-MEMS magnetic field sensors (see, e.g. STMicroelectronics, 2012); clearly, the integration of magnetometers in a MEMS process would be very advantageous both from a production cost perspective and for the further downsizing of the motion sensor. Demonstrations of such devices have been given in the scientific literature (Emmerich *et al.*, 2000) and it is thought that all-MEMS 9-DOF IMUs will soon enter the market.

In the following section, details will be given regarding the structure and working principle of MEMS accelerometers, gyroscopes and magnetometers; consideration will be given to various topologies and the focus will be on their specifications for consumer and automotive applications.

The operation of all of the MEMS devices described is governed by means of suitable electronics, generally embedded in an application-specific integrated circuit (ASIC). Even if the production of the MEMS and the integrated circuit on the same ASIC is possible (Geen *et al.*, 2002), the use of separate processes allows a better optimization of their respective performances, albeit at the cost of a larger area. The waste of area required by the combination of the MEMS and the ASICs will be considerably reduced with the advent of smart interconnections like through silicon vias (TSV) (Fischer *et al.*, 2011).

For a single device, the ASIC not only transduces the information regarding the quantity to be measured into an electrical signal and operates its conversion into the digital domain, but also provides an active function in the transduction principle (for instance, it keeps the device in oscillation for vibratory sensors, as will be detailed in the following section). The important point is that the choice of the driving and readout architecture of a MEMS device has a strong interdependence on the mechanical architecture, and both have an impact on the performance of a device – typically described in terms of sensitivity, resolution and operation bandwidth. When deriving the mathematical equations for these parameters, one should take into account these considerations to draw sound and valid conclusions.

In the next section, attention will be given to capacitive motion sensors (i.e. to devices where a change in the quantity to be measured results in a change of a suitably designed capacitance). The electronic readout then turns into a capacitance readout. This is the solution mostly widely used by MEMS sellers as it relies on a simple process that does not involve piezoelectric, piezoresistive or magnetic materials.

As a consequence, the *mechanical sensitivity* is defined here for every device as the capacitance variation per unitary variation of the quantity to be measured. *Electronic sensitivity* takes into account the readout circuit and will be expressed in terms of voltage variation at the output per unitary variation of the quantity to be measured. Other important parameters that must be taken into consideration during the project of a device are: the minimum measurable signal, or *resolution*, which corresponds to the overall system noise and is expressed in the same unit of the quantity to be sensed; the *full-scale*, corresponding to the maximum measurable variation; and, finally, the operation *bandwidth*, which corresponds to the maximum frequency of interest during the measurement.

Table 13.1 Typical specifications of MEMS accelerometers and gyroscopes for consumer/automotive applications

	Accelerometer	Gyroscope	Magnetometer
Full-scale-range (FSR)	±2 to ±16 g	2000 dps[a]	800 μT
Bandwidth	100–200 Hz	50–100 Hz	10–50 Hz
Resolution	1 mg	100 mdps[a]	1 μT

[a] dps = degrees per second.
Note: The data reported for the magnetometers are foreseen values, based on typical application requirements

Table 13.1 summarizes the state-of-the-art specifications for consumer and automotive applications.

13.2 Technical description of MEMS motion sensors: MEMS accelerometer

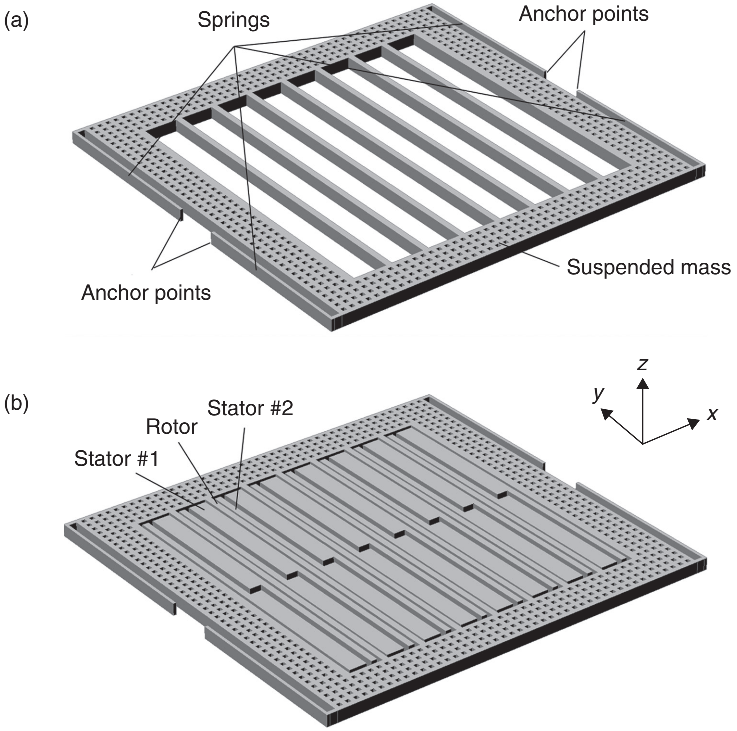
MEMS accelerometers are used to sense linear accelerations in the *x*, *y* and *z* directions. This section deals with the main issues faced by MEMS designers during the dimensioning of these devices. Figure 13.1 presents the schematic of a uniaxial MEMS accelerometer with capacitive readout (Biswas *et al*., 2007; Boser *et al*., 1996; Langfelder *et al*., 2011a). The device comprises a seismic mass, constrained to move only in the *x* direction by a set of springs anchored to the substrate. A set of capacitive parallel-plate differential cells is used to sense the displacements of the seismic mass. Each capacitive cell comprises a moving part (*rotor*) and two electrodes fixed to the substrate (*stators*).

13.2.1 Mechanical model

The 1-D movement of the seismic mass can be described by the well-known motion equation:

$$m\ddot{x} + b\dot{x} + k_m x = F_{\text{ext}} \qquad [13.1]$$

where m is the seismic mass value, b is the damping coefficient, k_m the mechanical elastic stiffness of the springs along the *x direction* and F_{ext} is the sum of external forces acting on the device. When designing MEMS devices, it is useful to study certain aspects regarding frequency. By transforming

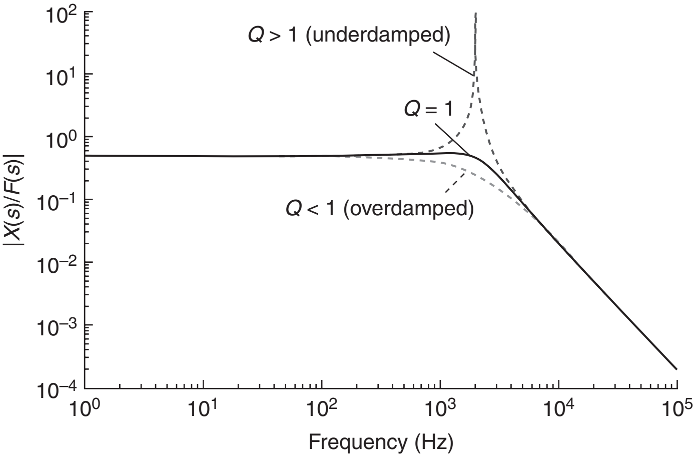


13.1 Schematic illustration of a uniaxial MEMS accelerometer. (a) View of the rectangular suspended mass enclosing the space for the sensing capacitors; (b) view of the full device with the differential stators.

Equation [13.1] by means of Laplace, the transfer function between an external force and the corresponding displacement can be written as:

$$\frac{X(s)}{F_{\text{ext}}(s)} = \frac{1}{m}\frac{1}{s^2 + (\omega_0/Q)s + \omega_0^2} \tag{13.2}$$

In Equation [13.2], $\omega_0 = \sqrt{k_m/m}$ is the mechanical resonance frequency of the device and $Q = \omega_0 m/b$ its quality factor. These two parameters are perhaps the most important to deal with when designing a MEMS accelerometer, as will be shown shortly. Figure 13.2 reports examples of the transfer function modulus of an accelerometer with an elastic stiffness of 0.5 N/m, a value in the typical range of consumer applications. The plot is characterized by a



13.2 Modulus of the transfer function between the external force and the displacement of a uniaxial MEMS accelerometer for different values of the quality factor.

flat response from the low frequency range until the resonance frequency is reached, governed by the relation:

$$\left|\frac{X(s)}{F_{\text{ext}}(s)}\right| = \frac{1}{k_m} \qquad [13.3]$$

Beyond the point at which resonance occurs, the frequency response decreases as a transfer function with two complex conjugate poles (–40 dB/decade). Therefore, the range $[0 < f < f_0]$ represents the mechanical bandwidth of the accelerometer, usually identified with the resonance frequency BW = f_0. For consumer and automotive applications, the frequencies of interest are lower than 100–200 Hz, so typically the device bandwidth is set around 1–2 kHz (one order of magnitude above). By substituting $F_{\text{ext}} = ma_{\text{ext}}$ in Equation [13.3], where a_{ext} is the external acceleration, we can have a better understanding of the importance of the resonance frequency value:

$$\left|\frac{X(s)}{A_{\text{ext}}(s)}\right| = \frac{m}{k_m} = \frac{1}{(2\pi f_0)^2} \qquad [13.4]$$

Equation [13.4] shows that, given an external acceleration, the displacement of the seismic mass depends on the ratio of the mass and the elastic stiffness, and not on their separate values alone. A common mistake is to assume that a larger mass leads to a more sensitive accelerometer; as shown in Equation

[13.4], this is ultimately not true. In conclusion, by setting the resonance frequency of the accelerometer (mainly considering the bandwidth requirements) one sets the relation between the displacement and the external acceleration. A large mechanical gain implies small bandwidths.

A further relevant issue can be observed in Fig. 13.2 in the correspondence of f_0, the modulus of the transfer function can either decrease smoothly or show a peak, depending on the quality factor value Q. The response is identified as over-damped where $Q < 1$, and as under-damped where $Q > 1$. This value depends greatly on the pressure set inside the package hosting the device during the process: capacitive accelerometers are packaged in such a range that the quality factor Q is typically lower than 1, in order to ensure an over-damped response from the device, without peaks in the transfer function. Indeed, a peak would result in large unwanted oscillations of the seismic mass in the event of external vibrations around f_0. Let us consider, for instance, mobile applications, where the accelerometer can be mounted adjacent to a loud speaker emitting a signal in the audio bandwidth (20 Hz to 20 kHz); or automotive applications, where the sensor operates in an acoustically harsh environment (Dean *et al.*, 2007): in both situations, any tones around f_0 would lead to a corrupted signal from the accelerometer.

13.2.2Differential capacitive sensing

Differential parallel-plate capacitive architecture represents the state-of-the-art of readout techniques currently integrated in most accelerometers for both the consumer and the automotive markets. The main advantage of this topology over other approaches (e.g. comb-fingers capacitance) is their unbeaten high-sensitivity in relation to the area occupation. This topology comprises an array of moving electrodes (*rotors*) attached to the suspended mass, and by two groups of fixed electrodes (*stators*), forming two MEMS capacitors C_1 and C_2 (see Fig. 13.1b) defined as:

$$C_1 = \frac{\varepsilon_0 A}{(x_0 + x)}; C_1 = \frac{\varepsilon_0 A}{(x_0 - x)} \qquad [13.5]$$

where ε_0 is the vacuum permittivity, A is the facing area of the plates, x_0 is the gap between a stator and a rotor in the rest position and x is the displacement of the rotor from its rest position. Often, in order to keep the area required for the device to a minimum, these capacitances are embedded as several differential capacitive cells in the central part of the device (see Fig. 13.1b). A positive displacement of the seismic mass (and thus of the rotors) in the x direction causes the decrease of C_1 and the increase of

C_2. Under the assumption of small displacements $x << x_0$, the differential capacitance variation can be computed as follows:

$$\Delta C = C_2 - C_1 = \frac{\varepsilon_0 A}{(x_0 - x)} - \frac{\varepsilon_0 A}{(x_0 + x)} \approx 2\frac{\varepsilon_0 A}{x_0^2}x = 2\frac{C_0}{x_0}x \qquad [13.6]$$

C_0 being the capacitance at rest between the stator and the rotor. The relation between the capacitance variation and the displacement can be thus approximated as linear for small displacements. A common approach is to design the device such that its dimensions – and thus the mechanical resonance frequency (see Equation [13.4]) and the capacitances – ensure the maximum accepted linearity error over the desired full-scale range (FSR). Typical values of the FSR are in the range ±2 to ±12 g. By substituting Equation [13.4] into [13.6], we can define what is identified here as the mechanical sensitivity, defined as the differential capacitance variation per variation of external acceleration:

$$\frac{\Delta C}{\Delta N_g} = 2\frac{C_0}{x_0}\frac{9.8}{(2\pi f_0)^2} \qquad [13.7]$$

where the external acceleration has been substituted by its corresponding number of *g*-units, $N_g = a_{\mathrm{ext}}/9.8\ \mathrm{m/s^2}$. The defined mechanical sensitivity is thus expressed in Farad per g-units (F/g).

13.2.3Thermo-mechanical noise

The minimum acceleration signal that can be measured is ultimately limited by noise sources, either intrinsic in the device or related to the readout electronics. With regard to intrinsic noise sources, the main contribution is associated with the thermo-mechanical fluctuations caused by the Brownian movements of the gas particles present inside the package hosting the device. Those particles hit the movable elements of the sensor causing a sort of stochastic *trembling*. The phenomenon is well known in the literature (Tsai *et al.*, 2005), where the noise power density of the force acting on the seismic mass as a consequence of this Brownian noise is reported as:

$$S_{Fn} = 4k_B Tb \qquad [13.8]$$

In Equation [13.8], k_B is the Boltzmann's constant and T is the absolute temperature. This spectral density has a frequency similar to that of white noise and it is expressed in $\mathrm{N^2/Hz}$. By using the relation between the external acceleration and the corresponding force, another important parameter characterizing the performance of a MEMS accelerometer can be obtained, the *acceleration noise power density*:

$$S_{gn} = \frac{S_{Fn}}{m^2 9.8^2} = \frac{4k_B T b}{m^2 9.8^2} \qquad [13.9]$$

expressed in g^2/Hz. The minimum detectable acceleration, or the sensor resolution, is defined as:

$$\sigma_g = \sqrt{S_g \mathrm{BW}} = \sqrt{S_g f_0} \qquad [13.10]$$

In general, the resolution required for the applications considered here is in the order of 1 mg. To give a numerical example, let us consider the following typical parameters for an accelerometer: a seismic mass $m = 5 \cdot 10^{-9}$ kg, a resonance frequency $f_0 = 2$ kHz and a quality factor $Q \sim 1$. We can evaluate the intrinsic resolution of the accelerometer as:

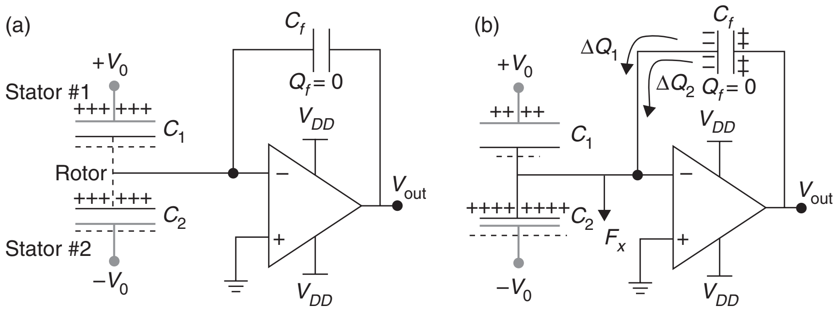
$$\sigma_g = \sqrt{\frac{4k_B T}{m^2 9.8^2} \frac{2\pi f_0 m}{Q} f_0} = \frac{f_0}{9.8}\sqrt{\frac{8k_B T \pi}{mQ}} \approx 3 \text{ mg} \qquad [13.11]$$

Note once more that a low f_0 has, in principle, a positive impact on the performance; nevertheless, apart from limiting the device bandwidth, the design of an accelerometer with a low f_0, obtained by decreasing the stiffness k, can lead to the well-known 'pull-in' phenomenon, as will be shown.

The relevant parameters for the design of a MEMS accelerometer have been identified and described in this section: the mechanical bandwidth, the sensitivity of the device and the acceleration noise density. The next step is to turn the variation in capacitance into an electrical signal from the MEMS.

13.2.4 Electronic readout

The electronic readout circuit converts the information on the variation in capacitance due to an external acceleration into a voltage signal. In this section, we will present the basic configuration for the readout of capacitive MEMS motion sensors: the charge preamplifier. Let us consider the electrical scheme represented in Fig. 13.3a, where the capacitive accelerometer is represented as its simplified electrical equivalent: a moving electrode between two fixed electrodes.



13.3 Schematic of the readout for a MEMS accelerometer. (a) The virtual ground of a charge amplifier is connected to the rotor of the accelerometer. The overall charge on the rotor is nominally null for $C_1(x = 0) = C_2(x = 0)$; (b) a displacement of the seismic mass determines a charge on the feedback capacitance which results in a variation of the output voltage.

The moving plate is connected to the virtual ground of the charge preamplifier, formed by an operational amplifier with a feedback capacitance C_f. The stators #1 and #2 are biased at a voltage $+V_0$ and $-V_0$, respectively. In this configuration the rotor is thus kept at a fixed voltage (ground, in this example), with a constant potential across the two capacitances. The expressions of the electrical charges on the two capacitances at rest can be written as:

$$Q_1 = C_1\left(V_0 - V_m\right) \qquad [13.12a]$$

$$Q_1 = C_2\left(V_0 - V_m\right) \qquad [13.12b]$$

and thus the net charge on the rotor node is null. Referring to Fig. 13.3b, let us suppose that the rotor moves in the x direction due to an external acceleration, causing a differential capacitance change. As a consequence, the charge on C_1 decreases while the charge on C_2 increases in order to satisfy Equation [13.12]. The charge variation on each capacitance with respect to the initial charge is simply:

$$\Delta Q_1 = -V_0 \Delta C_1 \qquad [13.13a]$$

$$\Delta Q_2 = +V_0 \Delta C_2 \qquad [13.13b]$$

These charges can be provided to the MEMS by the charge preamplifier through its feedback network, as shown in Fig. 13.3b. In particular, the total provided charge is the sum of the two charge variations. Considering the expression for the differential capacitance variation ΔC derived in Equation [13.6]:

$$\Delta Q_{\text{tot}} = \Delta Q_1 + \Delta Q_2 = V_0\left(\Delta C_2 - \Delta C_1\right) = V_0 \Delta C = 2V_0 \frac{C_0}{x_0} x \qquad [13.14]$$

This charge causes a variation of the output voltage of the preamplifier equal to:

$$V_{\text{out}} = \frac{\Delta Q_{\text{tot}}}{C_f} = 2V_0 \frac{C_0}{C_f} \frac{x}{x_0} = 2V_0 \frac{C_0}{C_f} \frac{N_g 9.8}{x_0 \left(2\pi f_0\right)^2} \qquad [13.15]$$

which give us the relationship between the preamplifier output voltage and the rotor (seismic mass) displacement. The factor $1/C_f$ can be considered to be the gain of the charge amplifier and should be chosen in order to have the maximum allowed output voltage change for an input number N_g of g-units corresponding to the FSR.

There is an important observation to be made at this point: the biasing voltages $\pm V_0$, essential for the capacitance readout are also sources of electrostatic forces acting on the seismic mass and altering its mechanical behavior with respect to what described in Section 13.2.1. Each stator indeed exerts an attractive force on the seismic mass that can be written as:

$$F_{el,1} = \frac{1}{2} \frac{dC_1}{dx} V_0^2 \qquad [13.16a]$$

$$F_{el,2} = \frac{1}{2} \frac{dC_2}{dx} V_0^2 \qquad [13.16b]$$

By summing these two forces, the total electrostatic force acting on the seismic mass – under the assumption of small displacements – can be obtained:

$$F_{el,\text{tot}} = F_{el,1} + F_{el,2} \cong 2 \frac{C_0}{x_0^2} V_0^2 x = K_e x \qquad [13.17]$$

Interestingly, it transpires that this electrostatic force results is proportional to the displacement x and it is therefore possible to define the term k_e which is generally referred to in the literature as the *electrical stiffness*. The mechanical behavior of the accelerometer when embedded in a readout design such as that presented in Fig. 13.3 can be therefore described by adding the electrostatic force term of Equation [13.17] in the motion equation:

$$m\ddot{x} + b\dot{x} + \left(k_m - k_e\right)x = F_{\text{ext}} \qquad [13.18]$$

The presence of the electrical term k_e has relevant effects on the behavior of the seismic mass. One of them is the pull-in phenomenon (Nielson *et al.*, 2006), a mechanical instability that arises when the electrical stiffness equals the mechanical stiffness. As detailed in the literature, apart from limiting the maximum signal bandwidth, the choice of a small value of k_m to increase the sensitivity by decreasing f_0 also leads to accelerometers with a small equivalent stiffness ($k_m - k_e$). In particular, in the limiting condition where $k_m = k_e$ the device loses its stable equilibrium point. The circuit presented is a simplified architecture of a typical electronic readout design. More complicated readout circuits, usually exploiting switching capacitor configurations and/or feedback circuits, can be used to decrease the power dissipated by the electronics and to solve the problems related to the pull-in instability (Langfelder *et al.*, 2011a; Seeger *et al.*, 2003).

The change in the effective stiffness of the system leads also to the definition of an equivalent resonance frequency:

$$f_{\text{eq}} = \frac{1}{2}\sqrt{\frac{\left(k_m - k_e\right)}{m}} \qquad [13.19]$$

It is worth noting that the effect of the electrical stiffness is always to diminish the total system stiffness, thus decreasing the resonance frequency. This effect has to be taken into account both for the choice of the bandwidth for a device and for the evaluation of the sensitivity, which can be obtained by substituting Equation [13.4] in [13.16] and by deriving with respect to external acceleration in g-units:

$$\frac{\Delta V_{\text{out}}}{\Delta N_g} = 2V_0 \frac{C_0}{C_f} \frac{9.8}{x_0\left(2\pi f_{\text{eq}}\right)^2} \qquad [13.20]$$

Equation [13.20] represents the *electronic sensitivity*, expressed in V/g. Usually, the device and the readout electronics are of dimensions such that the maximum voltage output allowed by the electronics $\Delta V_{\text{out,max}}$ is obtained with an acceleration signal corresponding to the target FSR of the accelerometer, $\Delta N_{g,\max}$.

The conclusions drawn here for in-plane accelerometers can be extended to out-of-plane (tilting) devices (Lemkin *et al.*, 1997; Selvakumar *et al.*, 1998) to implement a three-axis acceleration measurement unit.

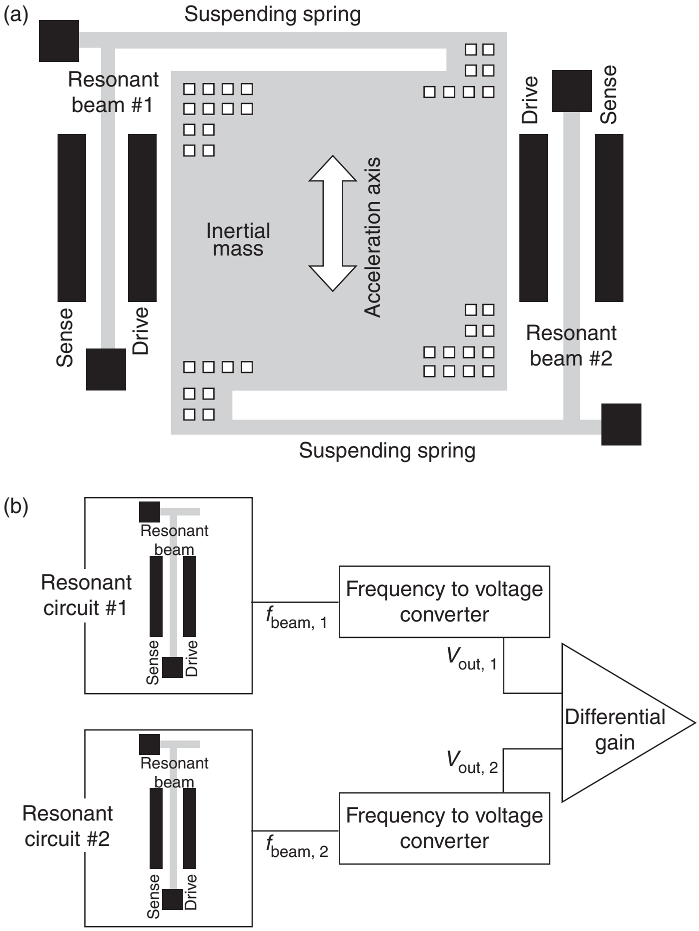
13.2.5Resonant accelerometers

The scalability of a capacitive accelerometer is somewhat limited by the pull-in phenomenon. Indeed, the bare scaling of the overall dimensions would lead to a lower mass and thus to a higher resonance frequency; in turn, leading to a lower sensitivity (see Equations [13.4] and [13.7]). To compensate for this change, one would design a device with a smaller mechanical stiffness k_m. This is, ultimately, in contrast with Equation [13.18], which states that a low mechanical stiffness can lead to an accelerometer with an overall null or negative equivalent stiffness ($k_m - k_e$), thus being mechanically unstable when biased through a readout circuit such the one described above.

One possible solution to overcome the pull-in problem, and one which has been given much attention in the scientific literature, is the resonant accelerometer (Aikele *et al.*, 2001; Roessig *et al.*, 1997; Seshia *et al.*, 2002). In this approach, an external acceleration acting on the MEMS device does not determine a change in a parallel-plate capacitance but, rather, a change in the resonance frequency of a suitably designed resonating element. A simple schematic of the working principle of such a device is shown in Fig. 13.4a. The device comprises an inertial mass suspended between a pair of suitably designed springs and two suspended beams, each having the same nominal resonance frequency at rest, $f_{\text{beam},1} = f_{\text{beam},2}$. For an external acceleration a_{ext}, the suspended mass is subject to the displacement given by Equation [13.4]. This displacement determines a differential stress on the beams, and thus a differential change in their resonance frequencies. It has been shown (Comi *et al.*, 2010) that for small stresses the differential frequency change $\Delta f_{\text{beam}} = (f_{\text{beam},1} - f_{\text{beam},2})$ can be linearized and it is in first approximation proportional to the external acceleration.

The readout electronics can be implemented as depicted in Fig. 13.4b: the beams are held in oscillation through suitable driving circuits (Tocchio *et al.*, 2012). The frequencies of each oscillator output are converted to voltage stages and their difference is computed through a final differential gain stage. An output signal proportional to the external acceleration is thus obtained.

Apart from the immunity to pull-in, which guarantees a high dynamic range, one of the advantages of this approach is that is takes up less room


(a)
Suspending spring
Resonant beam #1
Sense
Drive
Inertial mass
Acceleration axis
Drive
Sense
Resonant beam #2
Suspending spring
(b)
Resonant beam
Resonant circuit #1
Sense
Drive
$f_{beam, 1}$
Frequency to voltage converter
$V_{out, 1}$
Differential gain
$V_{out, 2}$
Resonant beam
Resonant circuit #2
Sense
Drive
$f_{beam, 2}$
Frequency to voltage converter

13.4 (a) Schematic structure of a resonant accelerometer: an acceleration acting on the inertial mass determines a differential stress on two beams, held in oscillation through capacitive driving and sensing electrodes (in black); (b) schematic representation of the readout electronics, including two oscillators, two frequency-to-voltage converters and a differential gain stage.

due to the small sensing elements (the two beams), unlike the array of stators required in the parallel-plate approach (see Fig. 13.1). The largest drawback is in the fact that a low pressure is required to guarantee a good quality factor for the resonating beams. A high-quality factor is mandatory prerequisite in order to minimize power dissipation within the two oscillating circuits. This low pressure requirement, however, conflicts with the need for the quality factor of the suspended mass to be kept low, in order to guarantee operation within the required bandwidth. This presently represents the main limitation to the deployment of this kind of device within the application considered in this chapter.

13.3 MEMS gyroscope

A gyroscope is a device used to measure the angular rate around a certain axis of rotation. Like other kinds of gyroscopes, MEMS gyroscopes also rely on the physical principle of the Coriolis force to perform the measurement. An object with a certain velocity v and an angular rate Ω around an axis orthogonal to the vector v, is subject to a Coriolis acceleration:

$$a_{\text{cor}} = 2\Omega \times v \qquad [13.21]$$

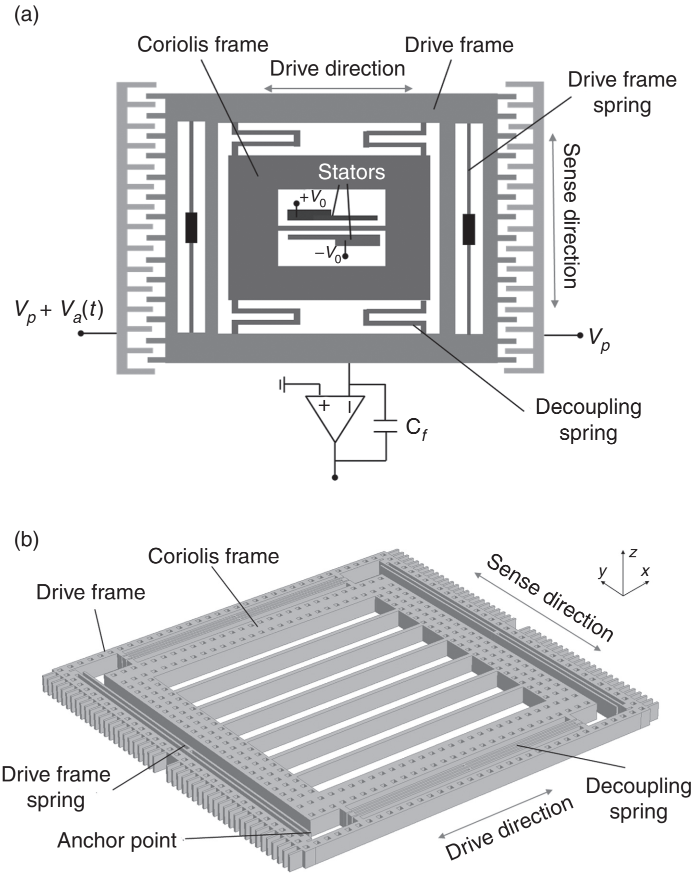
A corresponding Coriolis force acts on the object, with a direction orthogonal to the plane of both the axis of rotation and the direction of velocity, and with the following modulus:

$$F_{\text{cor}} = 2m\Omega v \qquad [13.22]$$

It is therefore important to ensure that the Coriolis force manifests only in presence of a velocity v; this immediately suggests that a MEMS gyroscope will also need to embed a driving section to apply a known velocity to a suspended mass. This is the first design difference with respect to the design of an accelerometer.

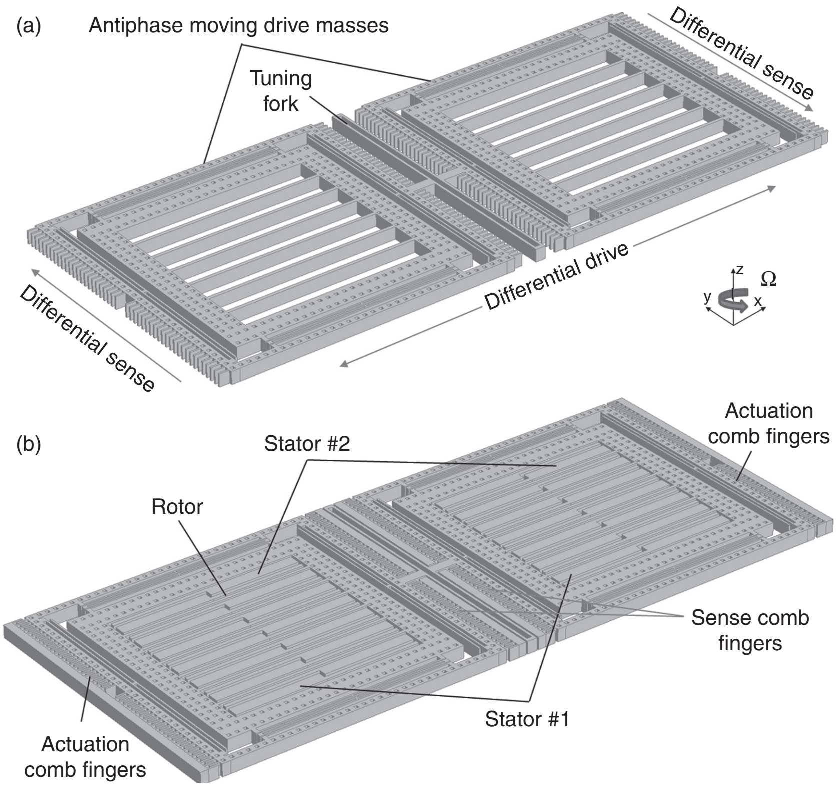
13.3.1 Working principle

An example of a MEMS gyroscope for the measurement of the angular rate along the z-axis is shown in Fig. 13.5 (Neul *et al.*, 2007; Sharma *et al.*, 2007). The device is formed by an external frame (the *drive* frame) suspended in such a way that it is free to move in the x-direction and is strongly constrained in the y-direction. Together with the fixed components, this frame forms two comb-finger capacitances that are used to apply the velocity v, as will be described. A set of suitably designed springs couples the drive frame to a second suspended frame (the *Coriolis* frame) in such a way that it is dragged by the drive frame along the x-direction but is also free to move in the y-direction. If a rotation around the z-axis occurs while the two masses are kept in oscillation with a velocity v, a Coriolis force pushes both the masses in the y-direction. The drive frame – constrained by the springs – does not move as a result of this force; the Coriolis frame is, instead, subject


(a)
Coriolis frame
Drive frame
Drive direction
Drive frame spring
Sense direction
Stators
$+V_0$
$-V_0$
$V_p + V_a(t)$
V_p
+
–
C_f
Decoupling spring
(b)
Coriolis frame
Drive frame
Sense direction
z
y
x
Drive frame spring
Anchor point
Decoupling spring
Drive direction

13.5 Illustration of a single mass gyroscope for sensing of angular rate along the z-axis. (a) Schematic top view with a simplified illustration of the voltages applied for driving and readout; (b) computer-aided design (CAD) 3-D view of the suspended mass showing the drive and Coriolis frame with the springs used to constrain the motion directions.

to a displacement which causes a differential capacitance variation with respect to the stators, designed in the same differential configuration as that seen for the accelerometers. The drive frame is usually kept in oscillation at the resonance frequency f_D of the drive mode. This allows the exploitation of the velocity amplification given by the quality factor Q_D in the drive direction, as described in the previous section. In order to obtain a large value of Q_D, the electrostatic actuation is undertaken using comb-finger actuators rather than parallel-plate cells: comb-finger actuators generally show a smaller damping coefficient and a larger linearity at high displacements.


(a)
Antiphase moving drive masses
Tuning fork
Differential sense
Differential drive
Differential sense
z
y
x
Ω
(b)
Stator #2
Rotor
Actuation comb fingers
Sense comb fingers
Stator #1
Actuation comb fingers

13.6 Illustration of a dual-mass tuning fork gyroscope. (a) View of the suspended frame with the tuning force which is used to mechanically constrain the antiphase drive mode; (b) view of the full device with the comb fingers for the drive oscillation in the *x* direction, and the differential stators used for capacitive sensing of the Coriolis force.

External accelerations in the *y*-direction can potentially disturb the measurement of the angular rate Ω as they can themselves determine a displacement of the Coriolis mass in the *y*-direction. To avoid this unwanted interference, the gyroscope can be designed in the dual-mass tuning fork configuration as reported in Fig. 13.6 (Geen *et al.*, 2002). In this implementation, the two drive frames are mechanically coupled by a central spring and are excited in the antiphase mode. As a consequence, in presence of a Coriolis force, the Coriolis frames move in the opposite direction. Conversely, acceleration has a common mode effect that can be cancelled out by adopting a differential readout. Further minimization of the signals caused by external accelerations can be obtained by filtering the acceleration signal, which typically occurs within a bandwidth <1 kHz, from the angular rate signal, which instead occurs around the drive modulation frequency f_D.

For simplicity, the equations that follow will be initially derived for a single mass gyroscope and finally combined by considering a differential capacitance readout such as that already described for the accelerometer.

13.3.2 Mechanical sensitivity

This section presents the main equations governing the operation of the MEMS gyroscope. In particular, we first consider the calculation of the velocity of the drive and Coriolis masses when suitable voltages are applied to the electrodes. The result is then used to calculate the Coriolis force and the obtained differential capacitance variation for a given angular rate.

As pointed out in the previous subsection, the drive frame and the Coriolis frame are kept in oscillation by applying appropriate voltages to the comb-finger capacitances as shown in Fig. 13.5. In the scientific literature, two main approaches are described: either (i) embedding the seismic masses as a resonating element inside a self-sustained oscillating circuit (Sharma *et al.*, 2007), or (ii) actuating the motion in an open-loop configuration by applying an appropriate AC voltage to the comb fingers. Figure 13.5 presents one version of a driving configuration: a DC voltage V_p is applied to both (right and left) the comb-finger stators and an AC voltage $v_a(t) = v_{a,\max} \sin(2\pi f_D t)$ is superimposed on only one of them; the seismic mass is considered to be kept at the ground voltage. In such a configuration, two electrostatic forces act on the drive and Coriolis frames:

$$F_1(t) = \frac{1}{2}\frac{dC_{D,r}}{dx}\left(Vp + v_a(t)\right)^2 \qquad [13.23]$$

$$F_2(t) = \frac{1}{2}\frac{dC_{D,l}}{dx}V_p^2 \qquad [13.24]$$

where the capacitances $C_{D,r}$ and $C_{D,l}$ are the capacitances formed by the comb-finger stators on the right and the left sides of the drive frame, respectively. These capacitances can be written as:

$$C_{D,r} = 2N_{\text{cell}}\frac{\varepsilon_0\left(L_{\text{ov}} - x\right)h}{x_0}; \qquad [13.25a]$$

$$C_{D,l} = 2N_{\text{cell}}\frac{\varepsilon_0\left(L_{\text{ov}} + x\right)h}{x_0} \qquad [13.25b]$$

In Equations [13.25a] and [13.25b], N_{cell} is the number of comb-finger cells, L_{ov} is the overlap length of two fingers, x_0 is the air gap in between them, x is the displacement of the drive and Coriolis frames along the x-axis and h is the out-of-plane thickness of the fingers. Under the assumption that the AC voltage amplitude is much lower than the DC voltage, $v_{D,\max}/2 << V_p$, the overall driving force acting on the two frames can be computed as:

$$F_D(t) = F_1 + F_2 = 2\frac{\varepsilon_0 N_{\text{cell}} h}{x_0} V_p v_a(t) \qquad [13.26]$$

One important aspect highlighted by Equation [13.26] is that the net electrostatic force applied by means of comb-finger capacitances is not a function of the displacement, unlike in parallel-plate capacitances. A force which is thus extremely linear with the value of the applied AC voltage $v_a(t)$ can be obtained even at large displacements. This is one of the main reasons why comb-finger actuation is preferred for the design of MEMS gyroscopes, where the sensitivity (as it will be shown shortly) is proportional to the maximum displacement of the Coriolis frame.

The motion of the drive and Coriolis frames along the x direction is governed by the same motion equation presented in Section 13.2 for the accelerometer, which can be now written as:

$$(m_D + m_{\text{cor}})\ddot{x} + b_D\dot{x} + K_D x = F_D(t) \qquad [13.27]$$

m_D and m_{Cor} are the mass values of the drive and Coriolis frames, respectively, b_D and k_D are the damping factor and the elastic stiffness characterizing the movement of the frames along the x-axis (in particular, k_D is related to the drive frame springs as represented in Fig. 13.5). Unlike accelerometers (which are operated at relatively high-pressures (e.g. ≈20–50 mbar) and are characterized by a poor quality factor), MEMS gyroscopes are operated at low pressures (e.g. ≈0.1–1 mbar) in order to achieve high Q factors, in the order of 1000. This feature allows the exploitation of the peak in the transfer function (see Fig. 13.2) to amplify the effects of the Coriolis force. Indeed, the driving voltage $v_a(t)$ is commonly a sinusoidal signal at a frequency equal to the resonance frequency of the drive mode. Therefore, the displacement in the x direction can be again written as a sine wave:

$$x(t) = F_D(t)\frac{Q_D}{k_D} = F_{D,\max}\sin(2\pi f_D t)\frac{Q_D}{k_D} \qquad [13.28]$$

The velocity is then computed as the time derivative of *x(t)*:

$$\frac{dx(t)}{dt}=\frac{Q_D}{k_D}2\pi f_D F_{D,\max}\cos(2\pi f_D t)=\frac{F_{D,\max}}{b_D}\cos(2\pi f_D t) \qquad [13.29]$$

where

$$F_{D,\max}=2\frac{\varepsilon N_{\text{cell}}h}{x_0}V_p v_{a,\max} \qquad [13.30]$$

is the maximum value of the driving force $F_D(t)$.

The modulus of the Coriolis force, acting on the Coriolis frame along the *y*-axis in presence of an angular velocity Ω (expressed in degrees per second, dps), can be computed through Equation [13.22] as:

$$F_c(t)=2m_{\text{cor}}\Omega\frac{2\pi}{360}\frac{dx(t)}{dt} \qquad [13.31]$$

Taking into account the fact that the drive frame is strongly constrained in the *y*-direction, the movement of the Coriolis frame along the *y*-axis – identified as the *sense mode* – is characterized by the following motion equation:

$$m_{\text{cor}}\ddot{y}+b_S\dot{y}+k_S y=F_c(t) \qquad [13.32]$$

in which b_S and k_S are the damping factor and the elastic stiffness along the *y*-axis. In particular, the value of k_S is related to the geometry of the decoupling springs as shown in Fig. 13.5. The corresponding transfer function between the displacement and the Coriolis force can be written as:

$$\frac{Y(S)}{F_c(S)}=\frac{1}{m_{\text{cor}}}\frac{1}{S^2+\omega_S/Q_S\,S+\omega_S^2} \qquad [13.33]$$

f_s and Q_s being the resonance frequency and the quality factor of the sense mode, respectively. The displacement of the Coriolis frame can be computed by multiplying the Coriolis force by the modulus of Equation [13.33], obtaining:

$$y(t)\frac{F_c(t)}{m_{\text{cor}}}\frac{1}{\sqrt{(4\pi^2 f_S^2-1)^2+\left(4\pi^2\left(f_S f_D/Q_S\right)\right)^2}} \qquad [13.34]$$

If we now consider the dual-mass tuning fork architecture shown in Fig. 13.6, Equation [13.34] describes the opposite displacements of each of the two Coriolis frames depicted. Using a suitably designed array of differential parallel-plate capacitances, the capacitive variations of these two frames can be summed up to double the output signal while simultaneously filtering the possible interference of external accelerations, as described. By combining Equations [13.6] and [13.34], the mechanical sensitivity of the gyroscope, defined as the sense capacitance variation per angular velocity variation (F/dps), can be seen to be:

$$\frac{\Delta C_D / \Delta\Omega = 2C_{0,S}/y_0\, y(t) = 2C_{0,S}/y_0\, F_C(t)}{m_{\mathrm{cor}}} \frac{1}{\sqrt{\left(4\pi^2 f_S^2 - 1\right)^2 + \left(4\pi^2 f_S f_D / Q_S\right)^2}} \qquad [13.35]$$

where $C_{0,S}$ and y_0 are the capacitance and the air gap at rest between the parallel-plate stators and rotors forming each of the sensing capacitances. It is worth noting that the sensitivity depends to a considerable degree on the value and the relative difference of the resonance frequencies f_D and f_S of the drive and sense modes. Under the assumption of perfectly matched frequencies, $f_D = f_S$, the sensitivity is maximized and can be written as:

$$\left.\frac{\Delta C_D}{\Delta\Omega}\right|_{\mathrm{matched}} = 2\frac{C_{0,S}}{y_0} F_C(t) \frac{Q_S}{k_S} = 4\frac{C_{0,S}}{y_0}\frac{2\pi}{360}\frac{m_{\mathrm{cor}} F_{D,\max}\cos(2\pi f_D t)}{b_D b_S 2\pi f_S} \qquad [13.36]$$

It is worth observing that the gyroscope sensitivity is proportional to the inverse of the resonance frequency value; therefore, low f_S (and f_D for perfect matching) values will ensure higher sensitivities. Nevertheless, MEMS gyroscopes, especially when tailored for consumer applications, are in general designed with an f_D and f_S of around 20 kHz in order to minimize possible interference caused by vibration in the acoustic bandwidth (Dean *et al.*, 2010).

Both the sensitivity and the signal bandwidth are a function of the frequency matching between the drive and the sense modes. Under the condition of matched frequencies, the bandwidth of the gyroscope corresponds to:

$$\mathrm{BW}_{\mathrm{matched}} = \frac{f_D}{2Q_S} \qquad [13.37]$$

For typical values of the frequency f_D (20 kHz) and the quality factor Q_D (1000), the signal bandwidth obtained would be only 10 Hz. Typical applications – both in the consumer and the automotive markets (Neul *et al.*, 2007), and even in the medical field (Della Santina, 2010) – require the measurement of angular rates on a larger bandwidth, around 50–100 Hz. In order to achieve such a large bandwidth, it is necessary to mismatch the frequencies of the drive and the sense mode by setting $f_S = f_D + \Delta f$: in this situation, the bandwidth of the gyroscope can be redefined as (Sharma *et al.*, 2008):

$$\mathrm{BW}_{\text{unmatched}} = \Delta f \qquad [13.38]$$

The mechanical sensitivity in the event of unmatched modes is significantly decreased since the quality factor of the sense mode is no longer fully exploited. In particular, for the range $f_s/2Q_s < \Delta f < f_s$ the mechanical sensitivity for the unmatched frequency condition can be expressed as:

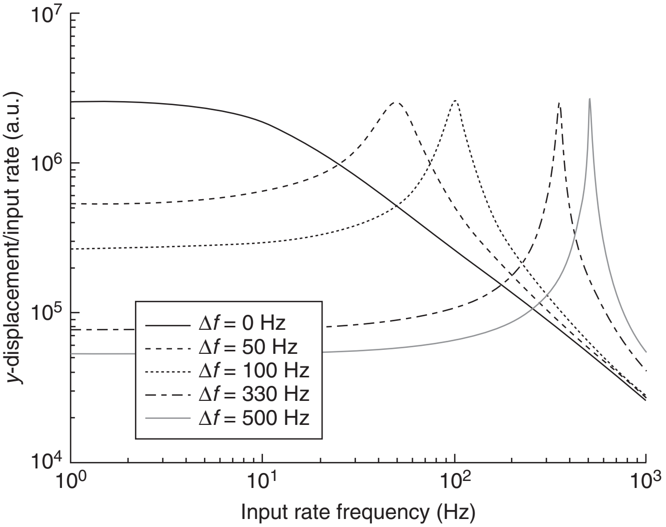
$$\left.\frac{\Delta C_D}{\Delta \Omega}\right|_{\text{unmatched}} = 2\frac{C_{0,S}}{y_0} F_C(t)\frac{Q_S}{k_S} = 4\frac{C_{0,S}}{y_0}\frac{2\pi}{360}\frac{F_{D,\max}}{b_D 2\pi f_S}\frac{\cos(2\pi f_D t)}{\mathrm{BW}_{\text{unmatched}}} \qquad [13.39]$$

As a rule of thumb, the higher the required bandwidth, the lower the mechanical sensitivity that can be obtained from a MEMS gyroscope. Figure 13.7 reports a set of representative simulations, showing the normalized transfer function along the y-axis as a function of the frequency of Ω, for different mismatches between the drive and the sense modes.

It is worthwhile noting that the design of perfectly matched frequencies, apart being limited to small bandwidth applications, is very challenging as it is easily affected by process variations. Random fluctuations in the experimental values of the masses and stiffness of the device result in different mismatches and thus lead to sensitivity variations between components.

13.3.3 Intrinsic device noise

Even if the gyroscope operates at a pressures ~1–2 orders of magnitude lower than the accelerometer, the intrinsic resolution of the device is also limited in this situation by the thermo-mechanical noise associated with the Brownian motion of gas particles inside the packaging. This noise source causes random movement in all the different frames that constitute the gyroscope. Considering the gyroscope architecture described in the previous section, both the movements of the drive frame and the Coriolis frames along the x-axis and the movement of the Coriolis frame along the y-axis are affected by thermo-mechanical noise. Nevertheless, the dominant term



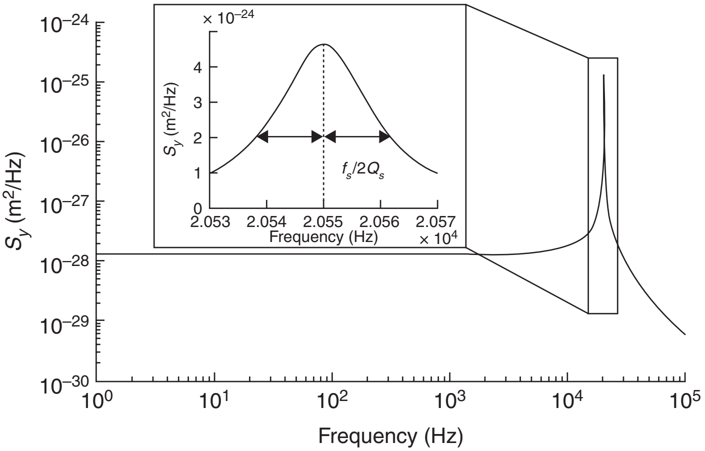
13.7 Transfer function of the Coriolis frame in the *y*-direction for the same angular rate as a function of the angular rate frequency. The different curves corresponds to different mismatches between the drive and sense modes. It can be observed that the larger the required bandwidth, the smaller the gain within the band.

is generally represented by the latter contribution, related to the Coriolis frames only. In particular, by using the definition of the noise power spectral density expressed in force (see Section 13.2), the noise power spectral density on the Coriolis frame, expressed in terms of displacement, can be computed as:

$$S_y = 4k_BTb_S\left|\frac{Y(s)}{F_c(s)}\right|^2 \qquad [13.40]$$

The native white noise contribution is therefore shaped by the transfer function of the sense mode, showing a peak around the resonance frequency f_s, as represented in Fig. 13.8. Under the assumption of matched frequencies, and assuming that an electronic low-pass filter with a bandwidth equal to $\mathrm{BW}_{\mathrm{matched}}$ is applied after a demodulation around f_S in the readout chain, the overall displacement noise power results:

$$\sigma_y^2 = 4k_BTb_S\left(\frac{Q_S}{k_S}\right)^2\frac{f_S}{2Q_S} \qquad [13.41]$$



13.8 Example of noise power spectral density in terms of displacement, as shaped by the sense mode. The close-up shows a detail of the peak, highlighting the full width at half maximum (FWHM) equal to f_s/Q_s. After demodulation, signals can be filtered with a low-pass bandwidth equal to $f_s/2Q_s$.

At this point, it is possible to evaluate the intrinsic resolution (i.e. the minimum detectable angular rate) of the device by computing the signal-to-noise ratio and equating it to unity:

$$\mathrm{SNR} = \frac{\left(2\,2\pi/360\; m_{\mathrm{cor}} F_{D,\max} / b_D b_S 2\pi f_S\, \Omega\right)^2}{4k_B T b_S\, Q_S f_S / 2k_S^2} \qquad [13.42]$$

$$\Omega_{\mathrm{rms}} = \frac{180}{F_{D,\max}} \frac{b_D}{b_S} 4\mathrm{BW}^2_{\mathrm{matched}} \sqrt{4k_B T \pi m_{\mathrm{cor}}} \qquad [13.43]$$

Regarding a dual-mass tuning fork gyroscope configuration, as noises are summed quadratically and signals are summed linearly, a further gain by a factor √2 in the signal-to-noise ratio (SNR) is obtained.

The considerations shown in this chapter can be easily extended to the design of gyroscopes for sensing the angular rate around the *x*- and *y*-axes (Prandi *et al.*, 2011; Trusov *et al.*, 2011).

13.3.4 References to resonant gyroscopes

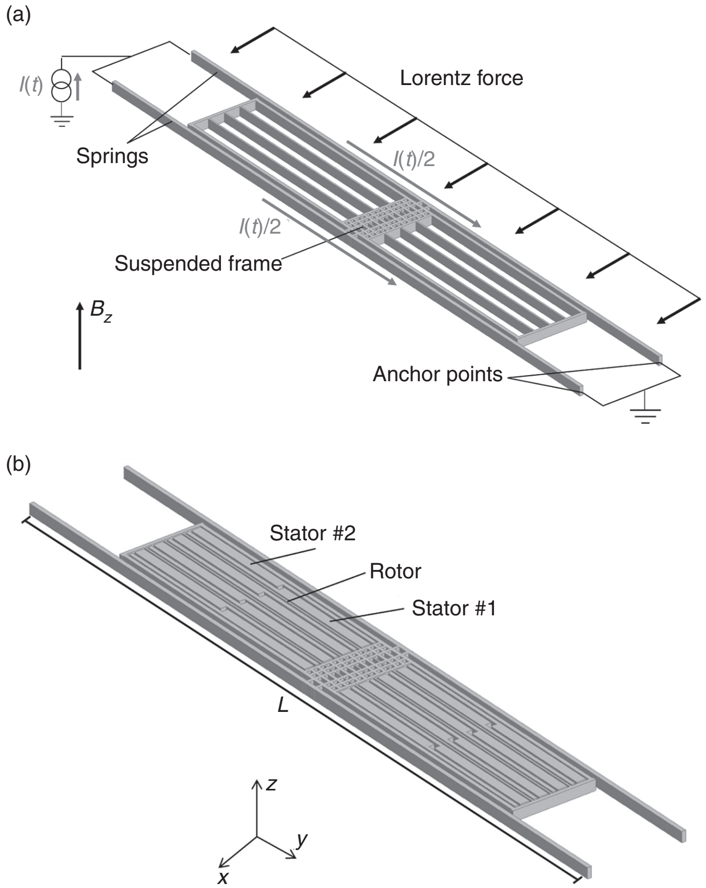
Resonant solutions have also been proposed for the gyroscope in order to augment the dynamic range and improve the resolution (see e.g. Seshia *et al.*, 2002b or Zotov *et al.*, 2012). In this kind of device, the driving section is similar to that described for the capacitive gyroscopes; that is, a suspended frame (or a pair of suspended frames) is kept in oscillation through a suitable driving circuit. The working principle for sensing is different from that of parallel-plate devices and is similar to that of resonant accelerometers (see Section 13.2.5): an angular rate results in a Coriolis acceleration, which determines a displacement that, in turn, causes a differential frequency shift of two resonating elements. The differential frequency shift is converted into a voltage difference that is thus proportional to the external angular rate.

13.4 MEMS magnetometer

There are many different methods by which to sense a magnetic field, depending on the intensity of the signal to be measured (Lenz, 2006). Magnetic fields can be as large as a few Tesla in MRI instrumentation; at the other end of the scale, the internal body activity manifests with minute magnetic field variations, as low as less than 1 nT. Therefore, given the large number of potential applications, there is an interest in developing miniaturized and low power magnetic field sensors. For the applications of interest in this chapter, considerate is important to bear in mind that, typically, the Earth's magnetic field has values between 20 and 70 μT; to guarantee a precision in the measurement of the field direction of about 1°, a resolution in the magnetic field measurement of ~0.5 μT per axis is required. Anisotropic magneto-resistance (AMR) magnetometers cope with this requirement and also with relatively low power operation (a few mW). State-of-the-art magnetometers for consumer application IMUs are based on this principle (STMicroelectronics, 2012). However, a complete IMU constructed in the same MEMS process would make devices very cheap and thus very attractive.

13.4.1 Working principle and mechanical sensitivity

Among the possible approaches for the implementation of a magnetometer in MEMS technology, one of the most promising is based on the Lorentz force principle with capacitive readout (Emmerich *et al.*, 2000; Kyynäräinen *et al.*, 2008; Thompson *et al.*, 2011a). This is because it does not require the integration of magnetic materials in the realization process (see Fig. 13.9). A pair of longitudinal springs of length L holds a suspended shuttle that, with suitable stators, forms a set of differential capacitors C_1, C_2, as described in the previous sections of this chapter. Suppose that a current $I(t)$ is induced in both springs. In the presence of a component B_z of the magnetic field in


(a)
I(t)
Lorentz force
Springs
I(t)/2
I(t)/2
Suspended frame
B_z
Anchor points
(b)
Stator #2
Rotor
Stator #1
L
z
y
x

13.9 Illustration of a z-axis MEMS magnetometer. (a) view of the suspended frame with a schematic of the working principle based on the Lorentz force arising in presence of a current *I*(*t*) and a magnetic field oriented in the *z* direction; (b) view of the full device with the differential stators for capacitive sensing.

the direction orthogonal to the plane of the figure, a force arises which has a total intensity:

$$|F_L(t)| = B_Z I(t) L \qquad [13.44]$$

The force is orthogonal to the plane of both B_z and $I(t)$, as depicted by the arrows in the figure. To give an initial idea of the intensity of this force, let us consider a spring length $L = 1$ mm (almost the maximum dimension reachable within a typical MEMS industrial package), a current $I(t) = 1$ mA (to cope with power dissipation, as will be clarified later) and the minimum field component to be detected $B_z = 0.5\ \mu\text{T}$. It transpires that the force has an intensity of 0.5 pN for the pair of springs, which is more than 2 orders of magnitude lower than the inertial forces as seen in Section 13.2. As a means with which to amplify the displacement obtained from such a force, the device can be operated at resonance, with a displacement amplification given by the quality factor (see Fig. 13.2). Let us suppose that the current has the following expression: $I(t) = I_0 \sin(2\pi f_0 t)$, where f_0 is the device resonance frequency. If the device is packaged at low pressure (like the gyroscope seen in the previous section), the displacement is amplified by the quality factor Q and is:

$$x(t) = \frac{|F_L|}{2k} \cdot Q = \frac{B_z \cdot I(t) \cdot L}{2k} \cdot Q \qquad [13.45]$$

(the factor 2 at the denominator is justified by the fact that the force is distributed along the springs and is not concentrated on the shuttle). This modulation of the current also has the advantage that it modulates the signal at a large frequency, possibly out of the $1/f$ noise of the readout electronics.

Consideration should be given here to the maximum Q that can be used; as the measurement bandwidth after the signal demodulation turns out to be equal to $\text{BW} = f_0/2Q$, in order to cope with typical required bandwidths for consumer and automotive applications (e.g. 50 Hz, see Table 13.1), the quality factor should not exceed a value which is 100 times lower than the resonance frequency.

By applying a differential capacitive readout technique, as seen in Section 13.2, the expression of the mechanical sensitivity (defined as the capacitance variation per unit of magnetic field change) can be derived, in the assumption of small displacements:

$$\frac{\Delta C}{\Delta B} = \frac{d}{dB_z}\left[2C_0 \frac{x(t)}{x_0}\right] = \frac{C_0}{x_0} \frac{I(t) \cdot L}{k} \cdot Q \qquad [13.46]$$

C_0 being the value at rest of each of the differential capacitors used for the readout. If we assume that each capacitor is formed by N cells, each having an area A_C, and if we write the expression of the quality factor in terms of the damping coefficient b, the following formula is obtained:

$$\frac{\Delta C}{\Delta B} = \frac{\varepsilon_0 A_C N}{x_0^2} \frac{I(t) \cdot L}{k} \cdot \frac{\sqrt{km}}{b} = \frac{\varepsilon_0 A_C N}{x_0^2} \frac{I(t) \cdot L}{b 2\pi f_0} \qquad [13.47]$$

It can be observed that, in principle, a low value of the resonance frequency f_0 would be preferable. However, due to the tiny values of the magnetic force, it is generally better to design the device with a frequency out of the audio bandwidth (i.e. above ~20 kHz) to avoid disturbances from amplified acoustic stimuli. As we have chosen f_0, it is now possible to determine the maximum quality factor to cope with the said measurement bandwidth requirements:

$$Q < \frac{f_0}{2\mathrm{BW}} \sim \frac{20\,\mathrm{kHZ}}{100\,\mathrm{HZ}} = 200 \qquad [13.48]$$

The quality factor is determined by the damping coefficient *b*. This parameter depends mainly on the geometry adopted for the design of the sensing cells. It is the squeezed film effect that dominates in the pressure regime that can be obtained with typical industrial packages (Langfelder *et al.*, 2011b). As a consequence, damping is set by the squeezing of the air between the moving plates and the stators. Therefore, a careful choice of the gap between the parallel plates and their area has to be done to cope with the requirement of Equation [13.48].

There are then practically only a few ways to increase, by means of design, the mechanical sensitivity of Lorentz force capacitive magnetometers: the maximization of the device length L; and increasing the driving current $I(t)$, which is however limited by power dissipation constraints. Great improvement can be obtained if the process itself is improved by establishing narrower air gaps x_0 between the parallel plates.

The description given here for the z-axis magnetometer can be easily extended to the other axes – for instance, by using tilting structures (Thompson *et al.*, 2011b) as previously mentioned with regard to the accelerometer and the gyroscope.

Bahreyni *et al.* (2007) also suggest the use of resonant sensing for magnetometers (as described for accelerometers and gyroscopes in sections 13.2.5 and 13.3.3).

13.4.2 Thermo-mechanical noise and intrinsic resolution

Let us now consider the effect of the thermo-mechanical noise power spectral density $S_{F,n}$, as defined in Section 13.2. For sake of simplicity, it will be assumed that the signal will be later filtered by the electronics with a bandwidth equal to the signal bandwidth defined above, BW = $f_0/2Q$ (see

Fig. 13.7). From Equation [13.8], the noise power, reported in terms of displacement, can be expressed as:

$$\sigma_x^2 = S_{F,n} \cdot \frac{Q^2}{k^2} \cdot \frac{f_0}{2} = S_{F,n} \cdot \frac{Q}{k^2} \cdot \frac{f_0}{2} \qquad [13.49]$$

Note that the random forces occur on the seismic mass and are not distributed to the springs. This explains the absence of a factor 2 at the denominator. The signal-to-noise ratio can be then written as:

$$\mathrm{SNR} = \frac{B_z IL/2kQ}{\sqrt{S_{F,n}\, Q/k^2\, f_0/2}} \qquad [13.50]$$

Finally, by putting the SNR equal to 1, the minimum measurable magnetic field can be derived as follows:

$$\left[\frac{B_z IL}{2k} Q\right]^2 = S_{F,n} \frac{Q}{k^2} \frac{f_0}{2} \qquad [13.51]$$

$$B_{z,\mathrm{rms}} = \frac{2}{IL}\sqrt{\frac{S_{F,n} f_0}{2Q}} = \frac{2}{IL}\sqrt{\frac{4k_B Tbf_0}{2Q}} = \frac{8\mathrm{BW}}{IL}\sqrt{k_B T \pi m} \qquad [13.52]$$

As an example, if we assume – as we have done so far – to have a driving current of 1 mA, a length of 1 mm, a resonance frequency of 20 kHz, a mass of 0.5 10^{-9} kg and a quality factor of 200, a minimum measurable magnetic field of ~1 μT is obtained, which shows that MEMS magnetometers based on the Lorentz force can be used effectively in IMUs for the applications considered in this section.

13.4.3 Effects of the readout electronics

It is worthwhile offering a brief discussion on the interdependence of the performances of the device and the readout electronics. Looking again at Equation [13.52], it is evident that the more current I is used to drive the device at resonance, the better the resolution. However, this implies a larger power dissipation by the Joule effect in the springs, $P_{\mathrm{MEMS}} = (I^2 \cdot R)$. The spring resistance R can be simply calculated from the material resistivity ρ and the spring section A_S:

$$R = \frac{\rho L}{A_S} \qquad [13.53]$$

A similar consideration can be made for the electronic noise $\sigma^2_{e\ln}$, which we have neglected so far: as a rule of thumb, the more the power is dissipated in the electronics P_{eln}, the better the noise performance. It can be concluded that, in the magnetometer design, the typical trade-off between resolution and power dissipation becomes a double trade-off, as summarized by the system below:

$$\begin{cases} \sigma^2_{\text{MEMS}} + \sigma^2_{e\ln} \le \sigma^2_{\text{application}} \\ P_{\text{MEMS}} + P_{e\ln} \le P_{\text{application}} \end{cases} \qquad [13.54]$$

This is caused by the fact that, in Lorentz-force-based magnetometers, there is an additional power contribution to power dissipation inside the device P_{MEMS} – which is not present, for instance, in accelerometers. The conclusion that can be drawn from this is that the interdependence of the components of a MEMS magnetometer and its readout electronics is even greater than for other devices. Optimization must be achieved at system level and not by separately considering the two sub-systems.

13.5 Conclusion and future trends

The aim of this section is to outline the possible evolution of smart IMUs over the next 5–10 years. It has been shown that accelerometers, gyroscopes and magnetometers can be designed using the same, relatively simple MEMS process and it is therefore expected that, in the very near future, (1–2 years) all-MEMS 9-DOF IMUs will be available. The reduced cost and dimensions of such IMUs will contribute to their widespread use in several fields of application. The next steps in the evolution of these devices can be identified at two different levels: hardware and software.

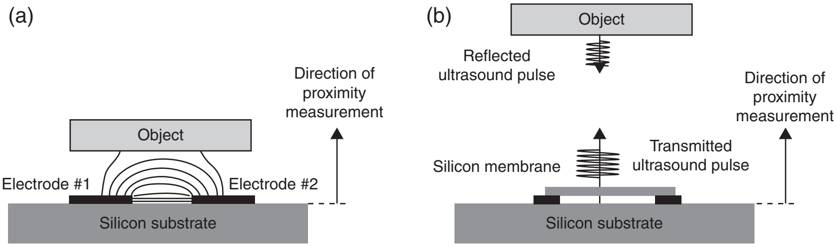
Innovation at the hardware level refers mainly to the development and integration of new devices, enabled by technological breakthroughs. As an example, another useful sensor that can be integrated into an all-MEMS IMU is the pressure sensor. This kind of sensor can be made using a suspended membrane (and not only a frame) and thus needs some modification to the technology (Zhang *et al.*, 2011). Indeed, in order to obtain relatively large sealed cavities made through a suspended membrane, it is necessary to change the process due to the fact that etching holes must be sealed after the release of movable parts. Sensing can be performed electrostatically through a bottom plate forming a capacitance the value of which changes with the external pressure. Such a sensor would add a precise and independent measurement of the height of the IMU (Analog Devices, 2011). This information could be exploited either for an initial calibration of the sensor position, or for further on-line corrections of the estimated position. It could also give information on the height in those environments where the Guide positioning systems (GPS) signal is easily lost: an example is the identification of the height (or the floor) inside a building or a skyscraper.

From the electronics and software point of view, it becomes clear that the larger the number of parameters that the system is able to measure, the more complicated the overall combination of the useful information becomes. Electronics would not only read out and digitize the information from the single units, but also process it to obtain more than the single, separate measurements (for instance, unwanted errors caused by an external acceleration on the measurements of other devices can be corrected *on-line* to augment precision, because the acceleration is measured within the same IMU). It is thus foreseen that, with the advent of such multi-function multi-parameter sensors, the software requirements will become more complex for the sensor fusion calculations. For instance, there will likely be a need for microcontrollers rather than the usual ASICs.

13.5.1 Combination of inertial measurement units (IMUs) with proximity sensors

Once the displacement and the motion can be accurately measured, further smartness is added to the system, if the surrounding environment can be recognized to some degree. Although infrared (IR) imaging, sensors can be used to recover the distance of an object (Jin *et al.*, 1998), their operation is typically limited to a low-light environment. Furthermore, IR sensors are generally made with materials other than silicon and thus they are inherently expensive.

The development of capacitive proximity sensors can overcome this issue, as they can be made through micromachining technologies similar to those used for the devices previously described in this chapter. In its simplest form, a proximity sensor is made through a set of co-planar electrodes which form a capacitance that is dominated by the fringe effects, as depicted in Fig. 13.10a (Chen *et al.*, 1998). As a consequence, an object that moves in the volume surrounding the electrodes changes the distribution of the field lines and thus determines a capacitance change (Langfelder *et al.*, 2012). The distance of the object moving above the electrodes can be measured if the object's material is known (a conductor or a dielectric will have different effects on the field lines; when considering a dielectric, the effect depends on its electrical permittivity, etc.). Typical measuring distances can range from a few μm to a few cm (Eun *et al.*, 2008; Lo *et al.*, 2012). As a consequence, this kind of proximity sensor is very useful in many industrial environments where space is very limited (microrobotics, machine automation, inspection tools, etc.).


(a)
Object
Electrode #1
Electrode #2
Silicon substrate
Direction of proximity measurement
(b)
Object
Reflected ultrasound pulse
Silicon membrane
Transmitted ultrasound pulse
Silicon substrate
Direction of proximity measurement

13.10 Illustration of a proximity sensor based on the changes of a fringe capacitance. (a) Schematic representation of a rangefinder based on a silicon membrane transmitting ultrasounds; and (b) detecting their echo.

Proximity sensors with a larger and material-independent measuring distance can be implemented because of the recent development of capacitive micromachined ultrasound transducers (cMUT). This kind of device can be obtained – in a similar manner to the pressure sensor – by building a sealed chamber with a membrane at the top and an electrode at the bottom (Saarilahti *et al.*, 2002). The bottom electrode is used both to excite the membrane electrostatically, in order to generate an ultrasound wave (typical frequencies between fractions to tens of MHz), and to sense the membrane capacitance variation when this is hit by the reflected wave. An array of such ultrasound wave transceivers can be used to obtain a depth image of the surrounding environment with a range in the order of a metre. A smart sensor capable of precisely measuring its motion and simultaneously monitoring the surroundings can have an enormous number of applications in gesture recognition, industrial automation, platform stabilization, medical equipment and robotics.

The scenario presented in this concluding section can be completed or partially modified by new evolutions in the industrial technologies for the mass production of MEMS devices. The integration of piezoelectric or piezoresistive materials (Guedes *et al.*, 2011), the integration of nano-wires within a MEMS process (Whalter *et al.*, 2012), the use of resonant sensors (Comi *et al.*, 2011) and the use of device stacking through TSV (Fischer *et al.*, 2011) are some examples of possible trends that can lead to further miniaturization and a higher degree of precision in sensing.

13.6 References

Aikele, M., Bauer, K., Ficker, W., Neubauer, F., Prechtel, U., Schalk, J. and Seidel, H. (2001), ‘Resonant accelerometer with self-test’, *Sensors and Actuators A, Physics*, vol. **92**, no. 1, August, pp. 161–167.

Analog devices (2010), 'ADIS16375', Low-profile low-noise six–degrees-of-freedom inertial sensor, technical datasheet, http://www.analog.com.

Analog devices (2011), 'ADIS16407', Ten degrees-of-freedom inertial sensor, technical datasheet, http://www.analog.com.

Bahreyni, B. and Shafai, C. (2007), 'A Resonant micromachined magnetic field sensor', *Sensors Journal, IEEE*, vol. **7**, no. 9, September, pp. 1326–1334, doi: 10.1109/JSEN.2007.902945.

Biswas, K., Sen, S. and Dutta, P. K. (2007), Mems capacitive accelerometer, *Sensors Letters,* **5** 471–484.

Boser, B. E. and Howe, R. T. (1996), 'Surface micromachined accelerometers', *IEEE Journal of Solid-State Circuits,* vol. **31**, no. 3, March, pp. 366–375.

Chen, Z. and Luo, R. C. (1998), 'Design and implementation of capacitive proximity sensor using microelectromechanical systems technology', *Industrial Electronics, IEEE Transactions on*, vol. **45**, no. 6, December, pp. 886–894.

Comi, C., Corigliano, A., Langfelder, G., Longoni, A., Tocchio, A. and Simoni, B. (2010), A resonant microaccelerometer with high sensitivity operating in an oscillating circuit, *Journal of Microelectromechanical Systems*, vol. **19**, no. 5, October, pp. 1140–1152.

Dean, R. N., Castro, S., Flowers, T. G. T., Roth, G., Ahmed, A., Hodel, A. S., Grantham, B. E., Bittle, D. A. and Brunsch, J. P. (2011), A characterization of the performance of a MEMS gyroscope in acoustically harsh environments, *IEEE Transactions on Industrial Electronics*, vol. **58**, no. 7, July, pp. 2591–2596.

Dean, R. N., Flowers, G. T., Hodel, A. S., Roth, G., Castro, S., Zhou, R., Moreira, A., Ahmed, A., Rifki, R., Grantham, B. E., Bittle, D. and Brunsch, J. (2007), 'On the degradation of MEMS gyroscope performance in the presence of high power acoustic noise', *IEEE International Symposium on Industrial Electronics*, ISIE Vigo, Spain, 4–7 June, pp. 1435–1440.

Della Santina, C. (2010), 'A device to restore balance', *Scientific American*, vol. **302**, pp. 68–71.

Emmerich, H. and Schofthaler, M. (2000), ' Magnetic field measurements with a novel surface micromachined magnetic-field sensor', *IEEE Transactions on Electron Devices*, vol. **47**, no. 5, May, pp. 972–977.

Eun, D. S., Kong, D. Y., Yoo, H. J., Hong, Y. M., Jang, J. M., Kang, T. W., Yu, I. S. and Lee, J. H. (2008), 'Design and fabrication of a MEMS-based multi-sensor', *Nano/Micro Engineered and Molecular Systems, 2008. NEMS 2008. 3rd IEEE International Conference*, Sanya, Hainan Island, China pp. 588–591, 6–9 January 2008.

Fischer, A. C., Grange, M., Roxhed, N., Weerasekera, R., Pamunuwa, D., Stemme, G. and Niklaus, F. (2011), 'Wire-bonded throughsilicon vias with low capacitive substrate coupling', *Journal of Micromechanics and Microengineering*, vol. **21**, no. 8, p. 085035.

Geen, J. A., Sherman, S. J., Chang, J. F. and Lewis, S. R. (2002), 'Single-chip surface micromachined integrated gyroscopewith 50/h Allan deviation', *IEEE Journal of Solid-State Circuits,* vol. **37**, no. 12, December, pp. 1860–1866.

Guedes, A., Shelton, S., Przybyla, R., Izyumin, I., Boser, B. and Horsley, D. A. (2011), 'Aluminum nitride pMUT based on a flexurally-suspended membrane', *Solid-State Sensors, Actuators and Microsystems Conference TRANSDUCERS, 2011 16th International*, Beijing, China, pp. 2062–2065, 5–9 June 2011.

Jin, X. C., Ladabaum, I. and Khuri-Yakub, B.T. (1998), 'Surface micromachined capacitive ultrasonic immersion transducers', Micro Electro Mechanical Systems, 1998. MEMS '98 Proceedings, Eleventh Annual International Workshop on, Heidelberg, Germany, pp. 649–654, 25–29 January 1998.

Kyynäräinen, J., Saarilahti, J., Kattelus, H., Kärkkäinen, A., Meinander, T., Oja, A., Pekko, P., Seppä, H., Suhonen, M., Kuisma, H., Ruotsalainen, S. and Tilli M. (2008), 'A 3D micromechanical compass', *Sensors and Actuators A: Physical*, vol. **142**, no. 2, pp. 561–568, 10 April 2008.

Langfelder, G., Frizzi, T., Longoni, A., Tocchio, A., Manelli, D. and Lasalndra, E. (2011), 'Readout of MEMS capacitive sensors beyond the condition of pull-in instability', *Sensors and Actuators A: Physical*, vol. **167**, no. 2, pp. 374–384.

Langfelder, G., Dellea, S., Zaraga, F., Cucchi, D. and Azpeitia Urquia, M. (2012), 'The dependence of fatigue in microelectromechanical systems from the environment and the industrial packaging', *IEEE Transactions on Industrial Electronics*, vol. **59**, no. 12, December, pp. 4938–4948 10.1109/TIE.2011.2151824, ISSN: 0278-0046.

Langfelder, G. and Tocchio, A. (2012), 'Differential fringe-field MEMS accelerometer', *Electron Devices, IEEE Transactions*, vol. **59**, no. 2, February, pp. 485–490.

Lemkin, M. A., Ortiz, M. A., Wongkomet, N., Boser, B. E. and Smith, J. H. (1997), 'A 3-axis surface micromachined ΣΔ accelerometer', *Solid-State Circuits Conference, 1997. Digest of Technical Papers. 43rd ISSCC., 1997 IEEE International*, pp. 202–203, 457, 6–8 February 1997.

Lenz, J. (2006), Magnetic sensors and their applications, *IEEE Sensors Journal*, vol. **6**, no. 3, June, pp. 631–649.

Lo, P., Hong, C., Tseng, S., Yeh, J. and Fang, W. (2012), 'Implementation of vertical-integrated dual mode inductive-capacitive proximity sensor', *Micro Electro Mechanical Systems (MEMS), 2012 IEEE 25th International Conference*, Paris, France, pp. 640–643, 29 January–2 February 2012.

Neul, R., Gómez, U.-M., Kehr, K., Bauer, W., Classen, J., Döring, C., Esch, E., Götz, S., Hauer, J., Kuhlmann, B., Lang, C., Veith, M. and Willig, R. (2007), 'Micromachined angular rate sensors for automotive applications', *IEEE Sensors Journal,* vol. **7**, no. 2, February, pp. 302–309.

Nielson, G. N. and Barbastathis, G. (2006), 'Dynamic pull-in of parallel-plate and torsional electrostatic MEMS actuators', *Journal of Microelectromechanical Systems,* vol. **15**, no. 4, August, pp. 811–821.

Prandi, L., Caminada, C., Coronato, L., Cazzaniga, G., Biganzoli, F., Antonello, R. and Oboe, R. (2011), 'A low-power 3-axis digital-output MEMS gyroscope with single drive and multiplexed angular rate readout', *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, pp. 104–106, 20–24 February 2011.

Robin, L. and Perlmutter, M. (2011), 'IMU and high performance inertial MEMS', Yole Développement technical report, http://www.i-micronews.com/upload/Rapports/Yole_IMU&High_Perf_Inertial_MEMS_Report_September_2011_web.pdf.

Roessig, T. A., Howe, R., Pisano, T. A. and Smith, J. H. (1997), 'Surface micromachined resonant accelerometer', in *Proceedings of Transducers*, Chicago, IL, vol. **2**, June 16–19, 1997, pp. 859–862.

Saarilahti, J., Blomberg, M., Häärä, A. and Kattelus, H. (2002), 'A novel method for processing capacitive micromechanical ultrasonic transducers (cMUT)', *Ultrasonics Symposium, 2002. Proceedings 2002 IEEE*, vol. **2**, pp. 1071–1074, 8–11 October 2002.

Seeger, J. I. and Boser, B. E. (2003), 'Charge control of parallel-plate, electrostatic actuators and the tip-in instability', *Journal of Microelectromechanical Systems*, vol. **12**, no. 5, pp. 656–671.

Selvakumar, A. and Najafi, K. (1998), 'A high-sensitivity z-axis capacitive silicon microaccelerometer with a torsional suspension', *Journal of Microelectromechanical Systems,* vol. **7**, no. 2, June, pp. 192–200.

Seshia, A. A., Palaniapan, M., Roessig, T. A., Howe, R. T., Gooch, R. W., Shimert, T. R. and Montague, S. (2002), 'A vacuum packaged surface micromachined resonant accelerometer', *Journal of Microelectromechanical System*, vol. **11**, no. 6, December, pp. 784–793.

Seshia, A. A., Howe, R. T. and Montague, S. (2002), 'An integrated microelectromechanical resonant output gyroscope', *Micro Electro Mechanical Systems, 2002. 15th IEEE International Conference*, LAS VEGAS, NEVADA, USA, 20–24 January, pp. 722–726.

Sharma, A., Zaman, M. F., Zucher, M. and Ayazi, F. (2008), 'A 0.1°/HR bias drift electronically matched tuning fork microgyroscope', *Micro Electro Mechanical Systems, 2008. MEMS 2008. IEEE 21st International Conference*, pp. 6–9, 13–17 January 2008.

Sharma, A., Zaman M. F. and Ayazi, F. (2007), 'A 104-dB dynamic range transimpedance-based CMOS ASIC for tuning fork microgyroscopes', *IEEE Journal of Solid-State Circuits,* vol. **42**, no. 8, August, pp. 1790–1802.

ST Microelectronics (2011), 'LSM330DLC', 3D accelerometer 3D gyroscope, technical datasheet, http://www.st.com.

ST Microelectronics (2012), 'INEMO-M1', iNemo system-on-board, technical datasheet, http://www.st.com.

Thompson, M. J. and Horsley, D. A. (2011a), 'Parametrically amplified z-axis Lorentz force magnetometer', *Journal of Microelectromechanical Systems,* vol. **20**, no. 3, June, pp. 702–710.

Thompson, M. J. and Horsley, D. A. (2011b), 'Low power 3-axis Lorentz force navigation magnetometer', *Micro Electro Mechanical Systems (MEMS), 2011 IEEE 24th International Conference*, Cancun, Mexico, pp. 593–596, 23–27 January 2011.

Tocchio, A., Caspani, A. and Langfelder, G. (2012), 'Mechanical and electronic amplitude-limiting techniques in a MEMS resonant accelerometer', *IEEE Sensors Journal,* vol. **12**, no. 6, June, pp. 1719–1725.

Trusov, A. A., Prikhodko, I. P., Zotov, S. A. and Shkel, A. M. (2011), 'Low-dissipation silicon tuning fork gyroscopes for rate and whole angle measurements', *IEEE Sensors Journal,* vol. **11**, no. 11, November, pp. 2763–2770.

Tsai, J. M. and Fedder, G. K. (2005), 'Mechanical noise-limited CMOS-MEMS accelerometers', *Micro Electro Mechanical Systems, 2005. MEMS 2005. 18th IEEE International Conference*, pp. 630–633, 30 January–3 February 2005.

Whalter, A., Savoye, M., Jourdan, F., Renaux, P., Souchon, F., Robert, P., Le Blanc, N. Delorme, C., Gigan, C. and Lejuste, O. (2012), '3-axis gyroscope with Si nanogauge piezo-resistive detection', *IEEE MEMS 2012*, 29 January 2012–2 February 2012, Paris, France, pp. 480–483, Gen. 2012.

Zhang, Y., Howver, R., Gogoi, B. and Yazdi, N. (2011), 'A high-sensitive ultra-thin MEMS capacitive pressure sensor', *Solid-State Sensors, Actuators and Microsystems Conference (TRANSDUCERS), 2011 16th International*, pp. 112–115, 5–9 June 2011.

Zotov, S. A., Trusov, A. A. and Shkel, A. M. (2012), 'High-range angular rate sensor based on mechanical frequency modulation', *Journal of Microelectromechanical Systems,* vol. **21**, no. 2, April, pp. 398–405.

14

MEMS print heads for industrial printing

S. LEE and J. CHOI, Sungkyunkwan University
(SKKU), Republic of Korea

DOI: 10.1533/9780857099297.2.402

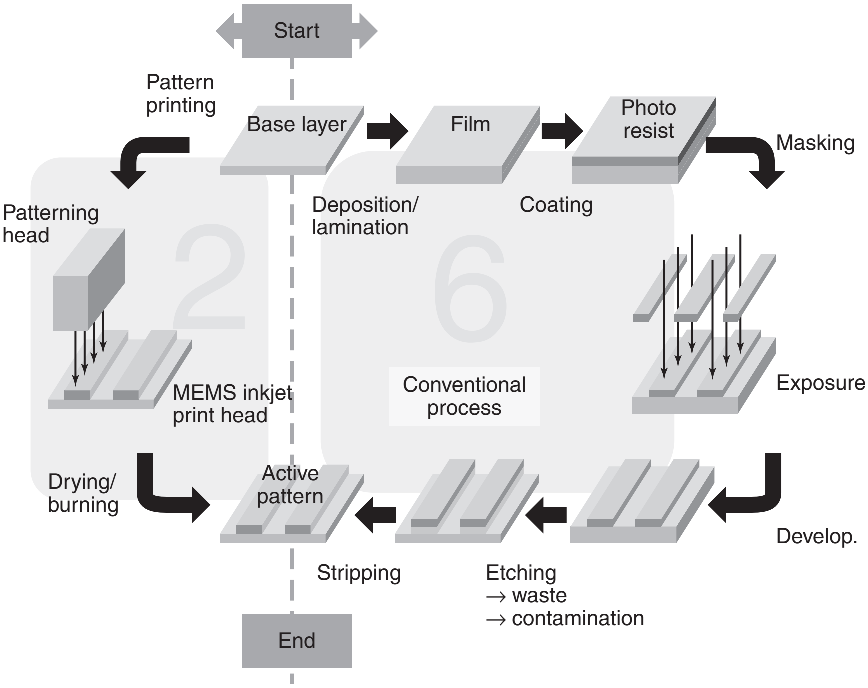
Abstract: With the recent prospect of micro electro mechanical systems (MEMS) print heads broadening their applications beyond digital printing for office use into industrial printing for the direct patterning of micro/nano devices and circuits, MEMS print heads have attracted renewed attention as a key player for realizing printed electronics. Cost-effectiveness, flexibility and environmental friendliness in fabrication are the main drivers of industrial printing. This chapter introduces MEMS print heads – especially, a new type referred to here as the electro hydro dynamics (EHD) print head. This has emerged recently as a viable and even better alternative to the conventional piezoelectric MEMS print heads for industrial printing, as it offers a higher resolution and thickness of patterning and can eject a high viscosity of ink. Nowadays, not only EHD droplet ejection, but also the EHD printing system for industrial applications is being investigated intensively with commercialization in mind. Furthermore, performance indices such as repeatability and stability are becoming a major issue to solve. This chapter also addresses briefly the effort to improve such reliability related indices.

Key words: electrohydrodynamic (EHD), print, MEMS, inkjet, ejection.

14.1 Introduction

In the past several decades, the generation and ejection of micro fluid droplets has been a subject of intensive investigation with its primary R&D objective being in the development of inkjet printing. Recently, there has been a tremendous surge of interest in diversifying the application of micro fluid droplet ejection in such a way as to meet a variety of physical, chemical, biological and engineering needs. For instance, the need for fine-printing or micro-dispensing of a sub-nano or a sub-pico litre volume of functional fluids has arisen due to the need for the cost-effective fabrication of high-sensitivity in sensing elements, for direct patterning of micro/nano devices and circuits and for carrying out large-scale combinatorial chemistry assays using expensive chemicals, to name but a few. The impending prospect that the technology for printing micro/nano fluid droplets is to serve as a key enabling technology for the coming printed electronics era can be seen in the heightened viability of directly fabricating sensors, printed circuit boards (PCBs), thin film transistors (TFTs), displays and biochips on flexible substrates. As the industrial application and commercial viability of micro/nano droplet ejection become more of a reality, performance measured by droplet size, jetting frequency, nozzle array density and the uniformity of ejected droplets become major concerns so as to satisfy the criteria for the particular industrial application under consideration (Lee *et al.*, 2008).

The technology of direct patterning based on an inkjet print head, possibly fabricated by MEMS technology, is now attracting attention due to its potential to replace the existing semiconductor fabrication process for many industrial applications. The conventional patterning based on semiconductor fabrication processing has to go through many steps for the formation of an active pattern because of the required deposition and etching processes with masking, as shown in Fig. 14.1. Direct patterning using an inkjet print head can radically reduce the processing steps, as illustrated in Fig. 14.1, since it can selectively and directly form active patterns on substrates, as if


Start
Pattern printing
Base layer
Film
Photo resist
Masking
Patterning head
Deposition/ lamination
Coating
2
6
MEMS inkjet print head
Conventional process
Exposure
Drying/ burning
Active pattern
Develop.
Stripping
Etching → waste → contamination
End

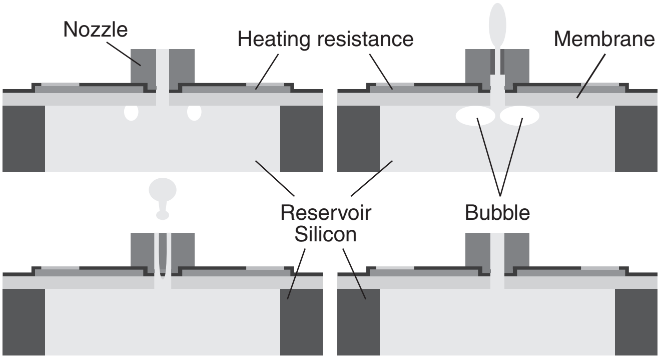
14.1 Comparison of a MEMS inkjet print head and conventional fabrication process.

writing letters or printing shapes directly on paper. As a result, processing costs are much lower than those of the conventional fabrication process due to the decrease in materials loss, the elimination of the need for masks and the shortening of processing time. Furthermore, it not only makes patterning large areas easier, but also provides an environmentally friendly process with a dramatic reduction in harmful materials from deposition and etching (Sirringhaus *et al.*, 2000; Singh *et al.*, 2010).

The advantages of the print head based direct patterning over the lithography-based traditional patterning have attracted significant attention from both industry and academia. In particular, the application of MEMS print heads to processes in industrial printing has drawn much attention to the development and commercialization of printed electronics.

As far as MEMS print heads are concerned, thermal-bubble and piezoelectric print heads have already achieved considerable commercial success in digital printing for office use (Bharathan and Yang, 1998; Le, 1998).

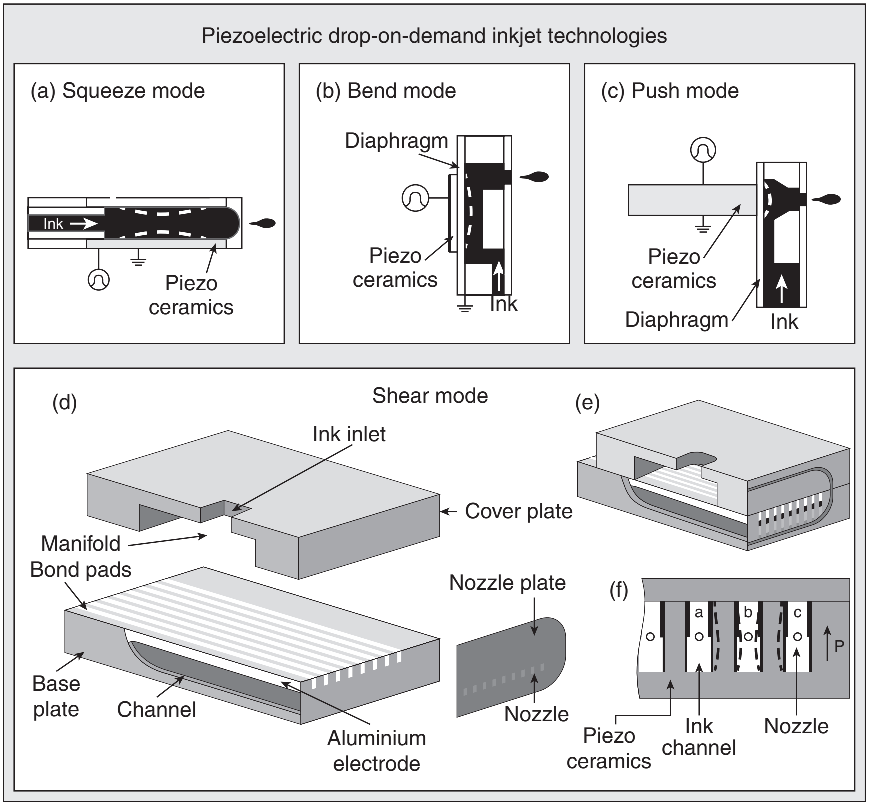
The thermal-bubble print head is currently the most successful print head in the office market. As shown in Fig. 14.2, when the ink supplied from the chamber is heated in the nozzle to several hundred degrees Celsius, bubbles are generated in the ink: the expansive force generated from the bubbles is used to push the ink through the nozzle, ejecting droplets of ink. The vacuum created as the bubbles draws replacement ink into the nozzle, generating a continuously forming supply of droplets (Allen *et al.*, 1985; Chen *et al.*, 1997). As the system is simple and the process is very easy and highly stable, it has been applied to Hewlett-Packard and Canon ink cartridges for inkjet printers (Calvert, 2001). However, the small and low-density nozzle array and ejection frequency required for thermal-bubble jetting incurs thermal problems when applied to industrial fabrication. Also, ink contains various solvents that can be severely compromised, as the thermal-bubble



14.2 The thermal-bubble print head. (*Source*: Zhou and Gué (2010), Copyright 2010, Elsevier.)

print head applies direct and considerable heat to the ink over a short time span (Ishida *et al.*, 2005; Choi *et al.*, 2008).

As heat levels cause serious problems in industrial printing for thermal-bubble print heads, research is being pursued on piezoelectric print heads. The piezoelectric print head (Fig. 14.3), first developed by Philips, pushes the ink from the chamber to the nozzle by means of the pressure caused by the vibration of the piezoelectric device situated behind the nozzle, thereby forming ink droplets. In particular, as the vibration of the piezoelectric device can be controlled by electrical signals, a piezoelectric print head can generate very tiny droplets by precisely adjusting small volumes of ink (Le, 1998; Wijshoff, 2010). Furthermore, as the ink is pushed by the pressure of the piezoelectric device vibration instead of using heat, the ink is not compromised. Moreover, it is possible to obtain a very high rate of ejection because of the ease with which low-voltage electrical signals can be input.


Piezoelectric drop-on-demand inkjet technologies
(a) Squeeze mode
Ink
Piezo ceramics
(b) Bend mode
Diaphragm
Piezo ceramics
Ink
(c) Push mode
Piezo ceramics
Diaphragm
Ink
(d)
Shear mode
(e)
Ink inlet
Cover plate
Manifold
Bond pads
Nozzle plate
Base plate
Channel
Aluminium electrode
Nozzle
(f)
a
b
c
P
Piezo ceramics
Ink channel
Nozzle

14.3 (a–f) The piezoelectric print head. (*Source*: Brünahl and Grishin (2002), Copyright© 2002, Elsevier).

For these reasons, Epson and other companies are now applying it to office printers (Le, 1998; Calvert, 2001).

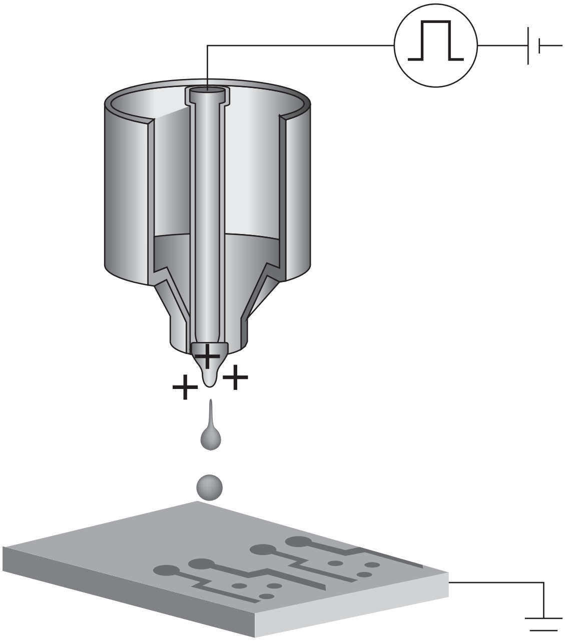
Therefore, the piezoelectric print head has been commercialized for industrial printing. The image of a representative piezoelectric printing system made by FUJIFILM Dimatix Inc. is available from URL: http://www.fujifilmusa.com/products/industrial_inkjet_ printheads/index.html.

The piezoelectric print head, however, cannot vary the size of droplets produced by a single nozzle; has also, there is a limit to the dimensions to which a nozzle can be reduced with regard to high-resolution patterns, as the size of the droplet is proportional to the size of the nozzle. Furthermore, reducing the nozzle size to create micro-droplets requires a reduction in the size of the piezoelectric device, leaving it unable to create sufficient force to eject the droplets. As nozzle density is difficult to increase, there is a limit in the degree of its integration. Moreover, there are many limitations when producing industrial ink because the pressure from the vibration of the piezoelectric device is insufficient to eject inks that are highly viscous (Ishida *et al.*, 2006; Byun *et al.*, 2008; Lee *et al.*, 2008).

As an alternative to the thermal-bubble and the piezoelectric print heads and their attendant shortcomings, EHD jetting, based on the direct manipulation of liquid by an electrical field, appears more promising. The EHD print head forms droplets by applying a high voltage, thus drawing ink to the nozzle, as shown in Fig. 14.4. In an electrical field, an electric charge is induced on the surface of the meniscus of the ink and electrical stress stretches the meniscus towards the direction of the field. An electrostatic field draws the meniscus of the liquid into a sharp cone, and charged liquid droplets are ejected from the nozzle when the electrostatic forces exceed the forces of the surface tension. Depending on the static biasing field, the duration and amplitude of the electric pulse and the physical properties of the liquid (such as surface tension, viscosity, electrical conductivity and dielectric constant), this device can generate a spray, a continuous stream or, under special conditions, discrete mono-disperse droplets (Kim *et al.*, 2007; Lee *et al.*, 2008).

The EHD print head causes no damage to the ink because no heat is generated – one of the major problems with thermal-bubble print heads. Also, the EHD print head can eject highly viscous liquids and increase the degree of integration without any of the fundamental limitations of a piezoelectric device. Furthermore, the EHD print head can make a range of droplet sizes that vary from several μm to several tens of μm in one nozzle because droplets can be formed only at end of the meniscus at the nozzle. Thus, ultra-high-resolution patterns (~10 μm) can be deposited on the substrate which it has been impossible to fabricate by any other printing means.

Moreover, the EHD print head can eject highly viscous ink because the ink is pulled by a strong electrostatic force created by the high voltage. Thus, the EHD printer has special characteristics: the ability to administer a wide



14.4 The electrohydrodynamic (EHD) print head.

range of inks and to form micro-/nano-sized patterns with high aspect ratios. Finally, array print heads can be manufactured easily because the structure of the EHD printer head is simple.

Therefore, the EHD print head is considered a core technology and has become a crucial fabrication process in the field of industrial applications.

14.2 Electro-hydro-dynamics (EHD) print head droplet ejection

EHD droplet ejection based on EHD spray phenomenon is known to be able to generate droplet by electric field and potential. EHD droplet ejection has different ejection modes; however, micro-dripping mode, pulsed cone-jet mode and continuous jet mode are the most appropriate ejection modes for generating desired droplet. Hence, EHD print head which can form droplet using electric field has various configurations for better droplet ejection results.

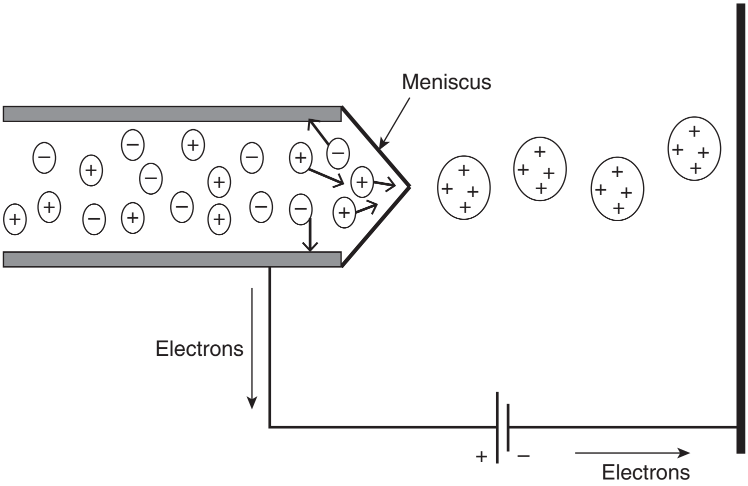
14.2.1 Principle of EHD droplet ejection

The EHD spray phenomenon, known as electrospray, has been known for a long time and related studies have been widely performed, mainly focused on the liquid atomization mode.

The EHD spray phenomenon is known to have been discovered by William Gilbert in 1600 (Mottelay, 1922). In the eighteenth century, Abbot Nollet experimentally showed the division of the main jet into very small jets through the break-up phenomenon by applying a strong electrical field to water. Also, Sir Rayleigh explained the instability of electrified droplets in 1879 (Eggers, 2008), and then Zeleny (1914, 1917) was the first to describe the possibility of droplet jetting by means of an electrical field, conducting an experiment to form droplets and taking pictures of this phenomenon. Since then, studies have continued and many results of droplet ejection have been published regarding various the settings and conditions of experiment.

The EHD print head, based on the EHD spray phenomenon which forms droplets through the interaction between electrostatics and fluidics, has the same principle as the EHD spray phenomenon that sprays liquid as tiny droplets by means of an electrostatic field. The basic principle of electrospraying is the electrical physical process. The charges in liquid and ink divided positive and negative ions with electrical characteristics due to the difference of electric potentials between the nozzle and the substrate. Thus, the electrical repulsive forces create the curved surface (meniscus) of the liquid. After that, the positive and negative ions are separated by the strong electrical force in the meniscus, and finally tiny droplets are formed at the furthermost part of the meniscus, as shown in Fig. 14.5 (Wilhelm, 2004).

Based on the EHD spray phenomenon, many studies have been conducted to form one droplet and to control the EHD droplet ejection. Several



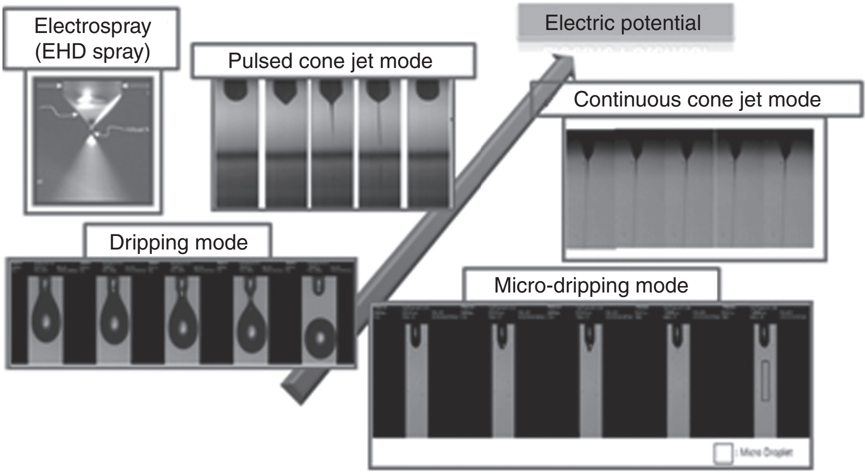
14.5 The schematic of EHD spray phenomenon.

researchers have attempted to manufacture an EHD print head that generates mono-disperse droplets, for the general purpose of MEMS print heads (Lee *et al.*, 2008). Thus, the results of EHD droplet ejection offer the basic foundation for the development of the EHD print head, and a wide range of experiments with droplet ejection have been conducted. Section 14.2.2 discusses this in greater detail.

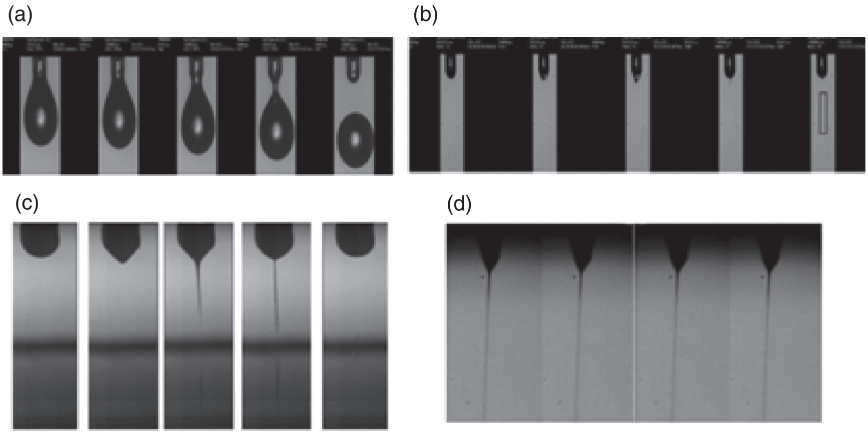
Many researchers have developed an EHD droplet ejection prediction and simulation model to demonstrate and explain EHD droplet ejection. The representative references (Smith, 1986; Ganan-Calvo, 1997; Notz and Basaran, 1999; Lastow and Balachandran, 2006; Collins *et al.*, 2007; Kim *et al.*, 2007; Li, 2007; Jaworek and Sobczyk, 2008) show various EHD droplet ejection modelling methods.

14.2.2 Various droplet ejection modes

The EHD spray phenomenon, explained in Section 14.2.1, was revealed to have various jetting modes according to the applied voltage and flow rate through a basic experiment for EHD droplet ejection (Cloupeau and Prunet-Foch, 1990). In particular, as shown in Fig. 14.6, the EHD spray phenomenon has different ejection modes depending on initial electric potential. However, ink properties were also revealed to be a major factor affecting ejection (Cloupeau and Prunet-Foch, 1994; Choi *et al.*, 2010). Thus, many studies are being undertaken to define the jetting mode based on ink properties. Also, research findings about droplet ejection pertaining to ink



14.6 EHD ejection mode by electric potential.



14.7 EHD ejection mode. (a) Dripping mode, (b) micro-dipping mode, (c) pulsed cone-jet mode and (d) cone-jet mode.

properties – such as surface tension, viscosity, electric conductivity and the dielectric constant – are helpful in the development of industrial EHD print head inks.

EHD droplet ejection modes can be typically divided into ten categories (Jaworek and Krupa, 1999). Among them, four modes – the dripping mode, micro-dripping mode, pulsed cone-jet mode and continuous cone-jet mode – are suitable for the EHD print head.

The dripping mode, the most basic ejection mode, is very similar to the falling of droplets with no electrical field. Droplets fall when the sum of the gravity and electrical force is greater than the surface tension, and the shape becomes spherical, as shown in Fig. 14.7a. However, with the dripping mode the pattern size is very large and the patterning process is too slow; this is due to the jetting frequency being too low and the droplet being larger than the outer diameter of the nozzle. Therefore, the dripping mode is inappropriate ejection mode for industrial use of an EHD print head.

The micro-dripping mode is similar to the dripping mode because it can form tiny spherical droplets; it is also similar to the pulsed cone-jet mode in terms of it forming long jet-like droplets (see Figs 14.7b and 14.7c). Compared with the dripping mode, however, the micro-dripping mode and the pulsed cone-jet mode are suitable for the EHD print head because they can form droplets which are smaller than the outer diameter of the nozzle and have a much faster jetting frequency.

In the micro-dripping mode and the pulsed cone-jet mode, the force of the electrical field at the end of the nozzle is sufficiently large, causing the meniscus to form a hemisphere or ellipse. Droplets which are much smaller than the outer diameter of the nozzle are formed only at the furthermost part of the meniscus, where the electrical field is concentrated. Therefore, the diameters of the droplets range from several μm to several hundred μm, and the size distribution of the droplets is generally even. Furthermore, the ejection frequency of the droplets ranges from several tens of Hz to several tens of kHz.

The continuous cone-jet mode is the basic method for EHD ejection. The jet formed by the meniscus is ejected continually without stopping and cutting, unlike in the pulsed cone-jet mode, as shown in Fig. 14.7d. Thus, frequency of ejection is almost infinite and the diameter of the jet is very small, so that when jet is printed on substrates, it can make sub-micron or micro patterns.

Therefore, the micro-dripping mode, pulsed cone-jet mode and continuous jet mode are certainly the most appropriate for industrial manufacturing processes.

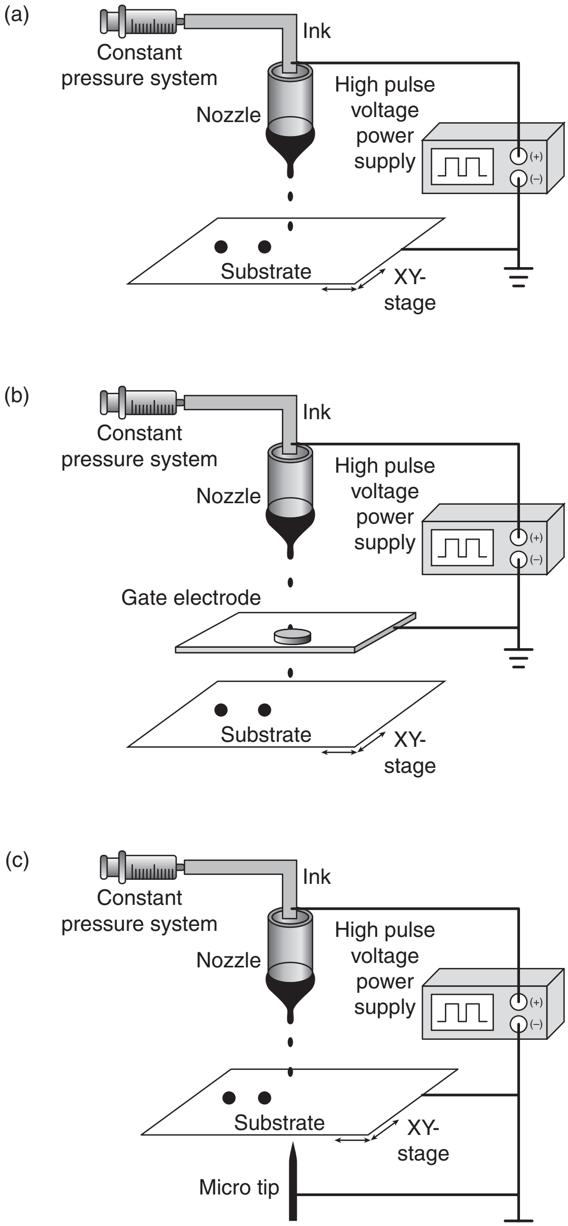
14.2.3 Configuration of the EHD print head

As with the conventional MEMS print heads (i.e., the thermal-bubble and piezoelectric print heads), the EHD print head basically consists of a nozzle, a chamber for storing ink and a pressure or flow regulator that supplies ink to the nozzle and controls its volume.

The EHD print head, as mentioned in Sections 14.2.1 and 14.2.2, forms droplets using the electrostatic field induced by the high voltage. Thus, EHD print heads are classified into three different types depending on how the electrical field to provide the high voltage is formed between the nozzle, substrate and electrode.

Figure 14.8a shows a direct EHD print head in which a high voltage supplier is connected to the nozzle and substrate, the droplets being formed by the difference of electric potential between the nozzle and the substrate. This has the same configuration as Fig. 14.5 which describes the basic principle of the EHD spray phenomenon in Section 14.2.1. The configuration is also identical to that used in electrospraying, which is the most commercialized application of EHD. The direct EHD print head can form droplets by making precise adjustments in the pressure and voltage, rather than by producing the EHD spray phenomenon using EHD spray equipment. This configuration is used in the simplest EHD print head and has been employed to set up the initial experiments on the EHD spray phenomenon. As it is easy to experiment with this printer head using sub-micron nozzles, high-resolution patterns using the direct EHD print head have been reported (Park *et al.*, 2007).

Figure 14.8b shows a gate electrode EHD print head in which a gate electrode is positioned between the nozzle and the substrate, and droplets are


(a)
Constant pressure system
Ink
Nozzle
High pulse voltage power supply
(+)
(−)
Substrate
XY-stage
(b)
Constant pressure system
Ink
Nozzle
High pulse voltage power supply
(+)
(−)
Gate electrode
Substrate
XY-stage
(c)
Constant pressure system
Ink
Nozzle
High pulse voltage power supply
(+)
(−)
Substrate
XY-stage
Micro tip

14.8 Three types of EHD printer head. (a) Direct EHD print head, (b) gate electrode EHD print head and (c) EHD print head with micro-tip.

formed by the difference of the electric potential between the nozzle and the gate electrode, with no electrical influence from the electrical field. When the jet (whether cone or pulsed cone) is ejected, however, most of the jet is divided into small droplets that are similarly sprayed after they pass through the electrode. Thus, studies are being conducted to solve this problem by adjusting the geometries of the electrode, substrate and applied voltage (Choi *et al.*, 2007).

Finally, Fig. 14.8c shows an EHD print head with a micro-tip configuration wherein a conical electrode (micro-tip) is positioned under the substrate. In this head, the conical electrode concentrates the electrical field on the meniscus, so the conical electrode is placed in line with the nozzle and the z axis. This has greater advantages for patterning because a more concentrated electrostatic force can apply to the meniscus compared with a direct EHD print head, as shown in Fig. 14.10a. If the substrate is dielectric, however, there is no benefit in concentrating the electrical field with a conical electrode.

Moreover, three different types of EHD printing system can be modified for better droplet ejection results (see Ishida *et al.*, 2006; Lee *et al.*, 2007).

14.3 EHD smart printing system

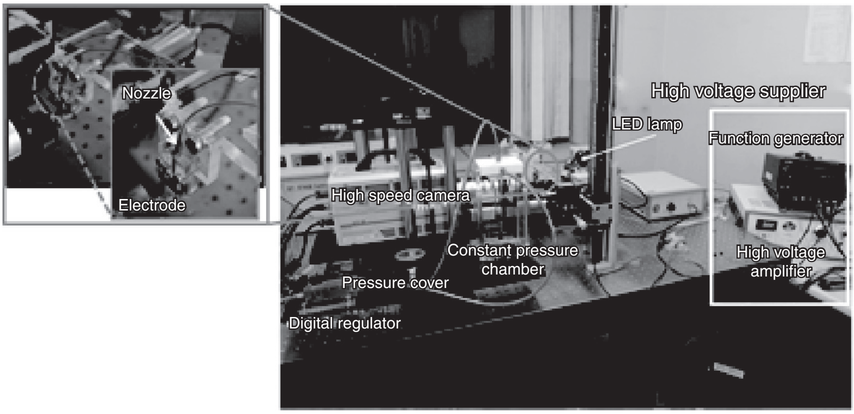
EHD printing system has recently attracted considerable attention for industrial manufacturing process. The section explains EHD printing system which can control the width and thickness of pattern and ensure repeatability and stability. Also, this section introduces EHD multi-head printing system for industrial applications.

14.3.1 EHD printing system

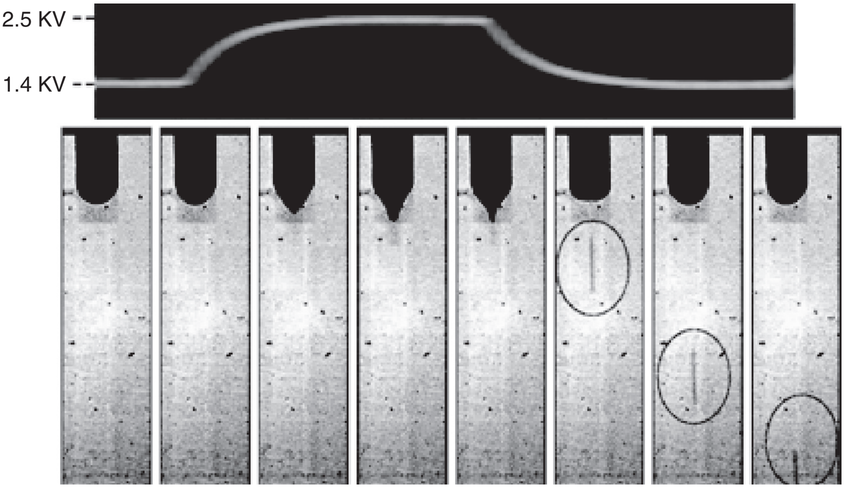
The EHD printing system requires a high voltage supplier (~3 kV) and means with which to supply voltage to the nozzle and the ink. Unlike the conventional thermal-bubble and piezoelectric MEMS printing systems, a droplet is formed from the meniscus in the nozzle by a strong electrical field. Figure 14.9 shows the prototype of the EHD printing system that connects the nozzle and the ink to the high voltage supplier.

Drop-on-demand (DOD) ejection is a key in industrial application. To achieve this, the high voltage supplier should generate an appropriately high voltage (~3kV) with a specific wave form rather than providing a constant voltage (i.e., a DC voltage). In other words, the high voltage supplier must be able to supply 2 kV or higher bias and pulse voltage waveforms;. this is essential to make the EHD printing system beneficial for industrial application. Thus, research into and production of a high voltage supplier with the relevant ejection parameters is as important as research on the EHD print head itself.

Using a high voltage supplier with bias and pulse voltage waveforms, the DOD ejection of droplets was examined for a DI water solution with SDS (Sodium Dodecyl Sulfate) (3 wt %) as shown in Fig. 14.10. A droplet is formed and ejected from the tip of the liquid meniscus in response to the pulse signal. With an applied external pulse of 500 Hz, the droplet is generated in accordance with the frequency. When a 1.1 kV pulse is applied based on a 1.4 kV bias voltage, the liquid meniscus is extracted from the tip of the nozzle and forms a cone shape. At the end of the pulse, the tiny droplet is broken from the apex of the meniscus and the remaining liquid is retracted. If the physical properties of the liquid (such as viscosity, surface tension, electric conductiv-


Nozzle
Electrode
High voltage supplier
LED lamp
Function generator
High speed camera
Constant pressure chamber
Pressure cover
High voltage amplifier
Digital regulator

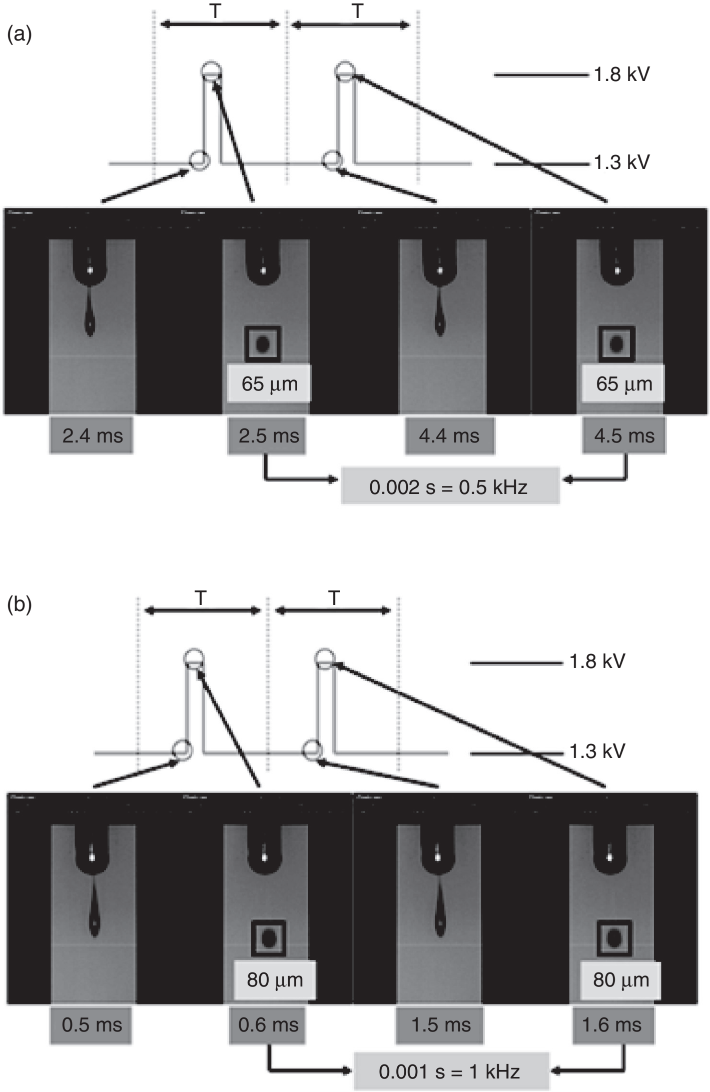
14.9 EHD printing system with high voltage supplier.



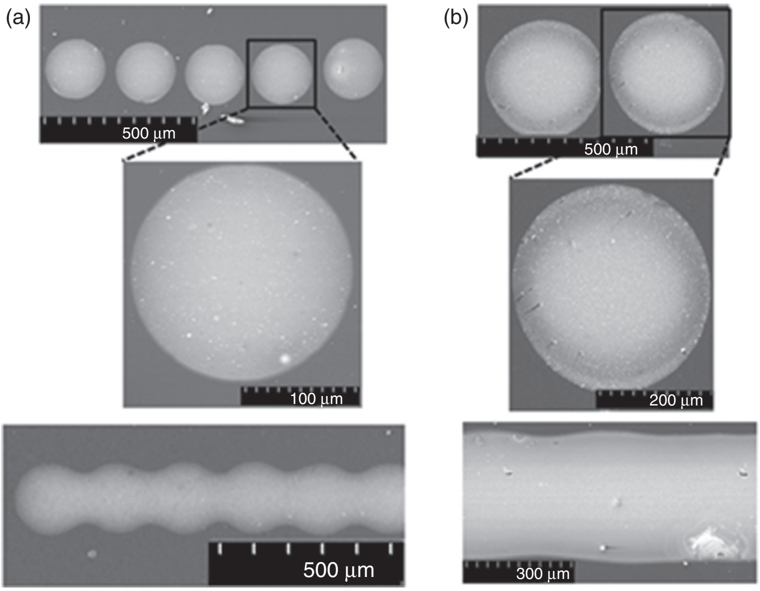
14.10 Drop-on-demand ejection characteristic using high voltage supplier at operating frequency of 500 Hz and duration of 0.2 ms. (*Source*: Lee *et al.* (2008), Copyright© 2008, Elsevier.)

ity and dielectric constant) are fixed and the optimal pulse and voltage are controllable, the DOD jetting system is feasible (Lee *et al*., 2008).

Based on a feasibility study, Choi *et al*. (2008) demonstrate the characteristic and patterning result of DOD ejection using a high voltage supplier with bias and pulse voltage waveforms. Figure 14.11 shows the DOD ejection characteristic as controlling applied bias and pulse voltage,


(a)
T
T
1.8 kV
1.3 kV
65 µm
65 µm
2.4 ms
2.5 ms
4.4 ms
4.5 ms
0.002 s = 0.5 kHz
(b)
T
T
1.8 kV
1.3 kV
80 µm
80 µm
0.5 ms
0.6 ms
1.5 ms
1.6 ms
0.001 s = 1 kHz

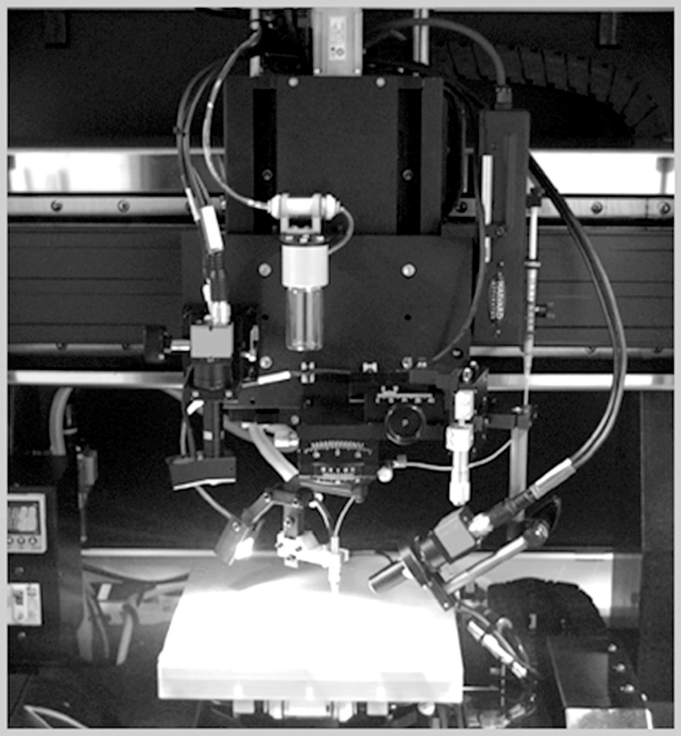
14.11 Jetting images of drop-on-demand results using conductive nanosilver ink: (a) 0.5 kHz and 5 μL/min and (b) 1 kHz and 15 μL/min. (*Source*: (Choi *et al*. (2008).)



14.12 SEM images of dot and line patterning results at drop-on-demand ejection: (a) 0.5 kHz and 5 μL/min, (b) 1 kHz and 15 μL/min. (*Source*: Choi *et al*. (2008).)

frequency, duty cycle and flow rate; Fig. 14.12 shows the scanning electron microscope (SEM) images of dots and lines produced with DOD ejection (Choi *et al*., 2008).

In order to realize an EHD printing system capable of performing various patterning processes, as with conventional printing systems the EHD system needs linear motors that can move the substrate and the EHD printing device in the direction of the *x, y* and *z* axes. Also required are: a vision system for the layered patterning and a control system that ensured compliance with the required print specifications. In particular, the system requires devices and programs to analyze the formation of droplets and patterns on the substrate through image processing and to control the transport speed and location of the substrate using information from the linear motor and high voltage supplier. Many research organizations and companies have recently developed EHD printing systems that satisfy these basic conditions and are selling them for EHD printing experiments and pattern formation. Figure 14.13 shows typically commercialized EHD printing systems.

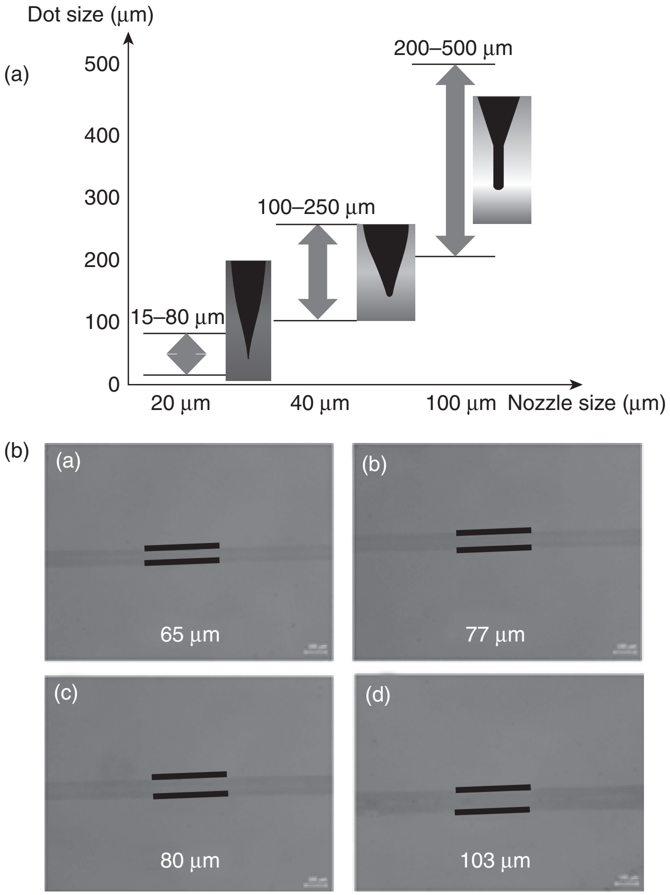


14.13 The commercialized EHD printing systems by ENJET Co. (*Source*: ENJET Co. brochure. Available from http://www.enjet.co.kr/index.html.)

14.3.2 Control of the pattern resolution and thickness

The parameters for controlling the pattern resolution and thickness of the EHD printing system are the nozzle/substrate (electrode) geometry, electrostatic field strength, ink properties, properties of the substrate surface, patterning repetition and motor velocity.

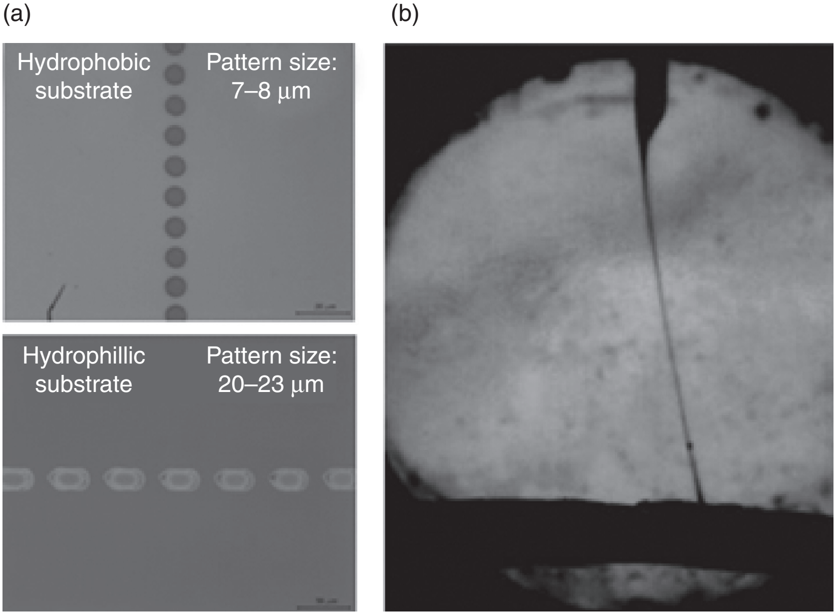
Even though the EHD printing system can eject droplets of various sizes from a single nozzle in response to the supplied voltage, it cannot actually jet sub-micron droplets and patterns in a nozzle with a diameter of 100 μm. Therefore, the nozzle should be smaller to enable sub-micron patterning, which is one of the strong points of the EHD printing system, as shown in Fig. 14.14a. Also, the pattern sizes of the EHD printing system are influenced considerably by the applied voltage. Increasing the applied voltage causes an increase in pattern volume, as shown in Fig. 14.14b. In addition,



14.14 The relationship between patterned dot size and nozzle size and (b) the result of patterning at applied voltage (a) 2.0 kV; (b) 2.2 kV; (c) 2.4 kV and (d) 2.8 kV.

pattern width and the scattering effect of pattern are affected by the distance between the nozzle and the substrate (or electrode), because the electrical field applied to the meniscus can be varied depending on the distance between the nozzle and the substrate (or electrode). Therefore, the design of the nozzle/substrate (electrode) geometry is important to achieving the desired patterning.

The EHD printing system also forms patterns as droplets are rapidly deposited on the substrate, as is the case with thermal-bubble and piezoelectric printing systems. Thus, the surface energy determines the degree to which the droplets spread on a substrate. Therefore, as shown in Fig. 14.15a,



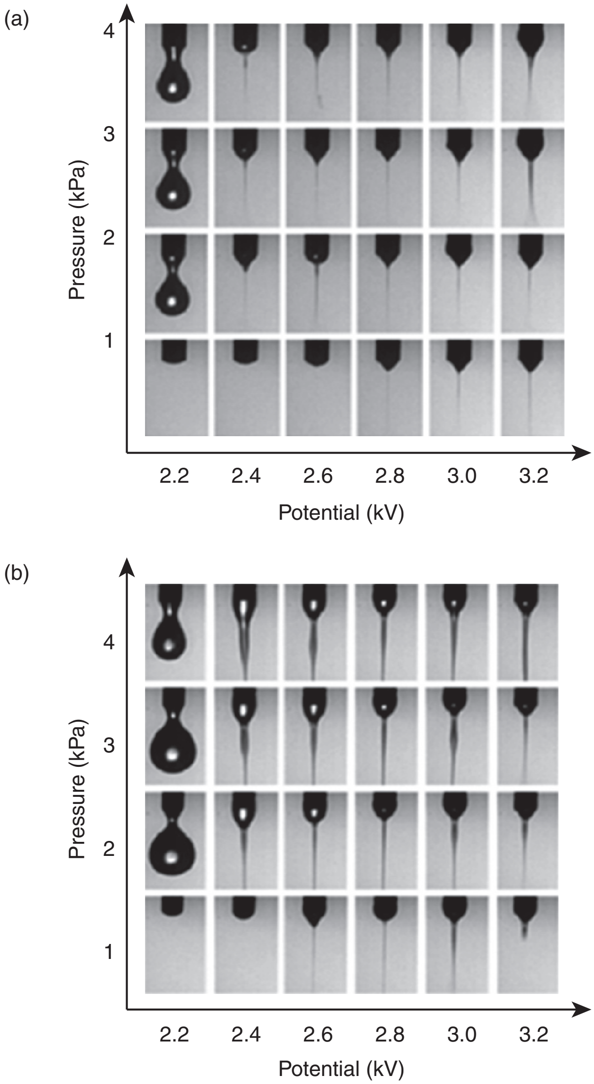
14.15 (a) Patterns on hydrophobic and hydrophilic substrates, and (b) the jetting image of ink with ~12 000 cPs. (*Source*: Jayasinghe and Edirisinghe (2004), Copyright© 2004, Elsevier.))

the pattern on a hydrophobic substrate is smaller and thicker than the pattern on a hydrophilic substrate. Also, through patterning repetition reprints on printed patterns not only can thick patterns be obtained, but the thickness of the patterns can also be controlled.

Finally, the EHD printing system has a larger range of ink viscosity. Therefore, the EHD printing system can eject inks of a higher viscosity and form thicker patterns with them than with conventional printing systems. Figure 14.15b shows the ejection of a highly viscous ink (approximately ~12,000 cPs) (Jayasinghe and Edirisinghe, 2004).

14.3.3 Repeatability and stability

The repeatability and stability of droplet ejection and patterning in the EHD printing system are important factors for applying the system to industrial manufacturing processes. To ensure repeatability and stability, the meniscus should not reach the outer nozzle and the movements should be fixed in the repeated ejection process. To establish the shape of the meniscus without modifying the nozzle, hydrophobic materials are applied externally to the nozzle. Figure 14.16 depicts stable ejection where the meniscus does not overflow on the outer part of the nozzle. The hydrophobic nozzle has the advantage of ejecting droplets by retaining the meniscus shape at the moment of ejection because of the relatively large contact angle of the

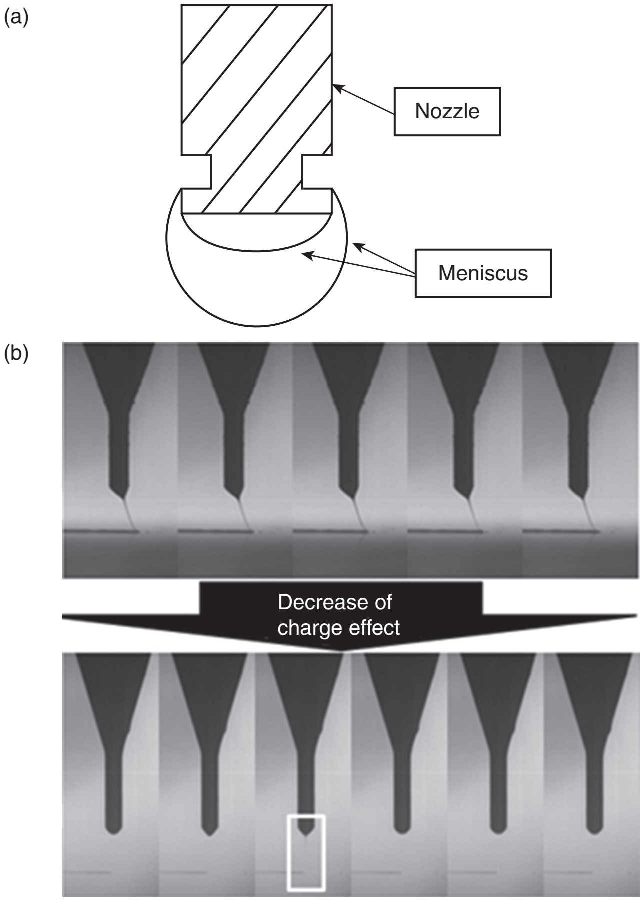


14.16 The moment of ejection and meniscus shape images: (a) hydrophobic nozzle and (b) hydrophilic nozzle.

hydrophobic nozzle surface, as shown in Figs 14.16a and 14.7b (Kim *et al.*, 2010a).

Figure 14.17a also shows that the EHD printing system can offer excellent performance regarding repeatability and stability without overflowing and damping at the meniscus on the nozzle. The nozzle includes a concave section formed along its outer circumference to ensure repeatability and stability (Kim *et al.*, 2010b).

The effects of various pulse voltages, frequencies, duty cycle and pressures for meniscus deformation and ejection were investigated to find the optimal region for EHD DOD printing. The research showed that the restoration



14.17 Improvements in the repeatability and stability of an EHD printing system: (a) a nozzle with a concave section and (b) droplet ejection and patterning as decreasing charge effect.

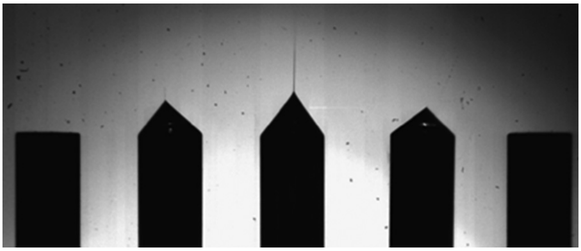
time of the meniscus applied to the pulse voltage is an important factor of EHD DOD printing regarding the relationship between the pulse voltage and the frequency (Son *et al.*, 2010).

Droplets themselves have electrical charges because they are formed by the electrostatic force in the EHD printing system. The electrical charges in the droplets cause the repulsive force between the substrate and the droplets. Thus, droplets cannot remain straight on the substrate and form a pattern at the exact place when the droplets are patterned on substrates, especially a dielectric substrate such as pure glass wafer. Therefore, many studies are being conducted to remove the electrical charge from a substrate; alternatively, the distances and applied conditions between the nozzle, substrate and electrode are controlled in order to reduce the charge effect. Figure 14.17b shows improved droplet ejection and patterning where the charge effect has been countered by controlling and optimizing the pulse voltage.

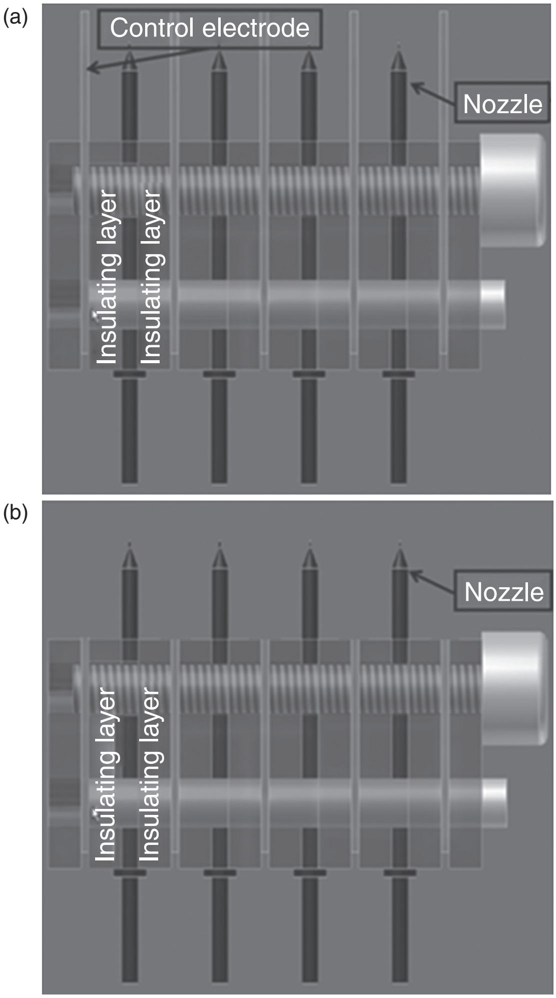
14.3.4 EHD multi-head printing system

An EHD printing system with multiple nozzles would shorten fabrication time (tack time) for industrial applications. Thus, the interference of the electrical field between the nozzles must be overcome for EHD multi-head manufacture. In other words, if the voltage supplied to form droplets influences the electrical field of the neighbouring nozzle, this nozzle will not form droplets due to its changed electrical field. In addition, accurate patterning at the desired location becomes impossible because of a bent jet or droplet, as shown in Fig. 14.18 (Si *et al.*, 2007).

Therefore, study is being undertaken into EHD multi-head printing systems that can reduce or remove the interference to the electrical field between the nozzles (Lee *et al.*, 2011). Two representative EHD multi-head printing systems are shown in Fig. 14.19a: the modified gate electrode EHD print head and (b) the direct EHD multi-head print head.



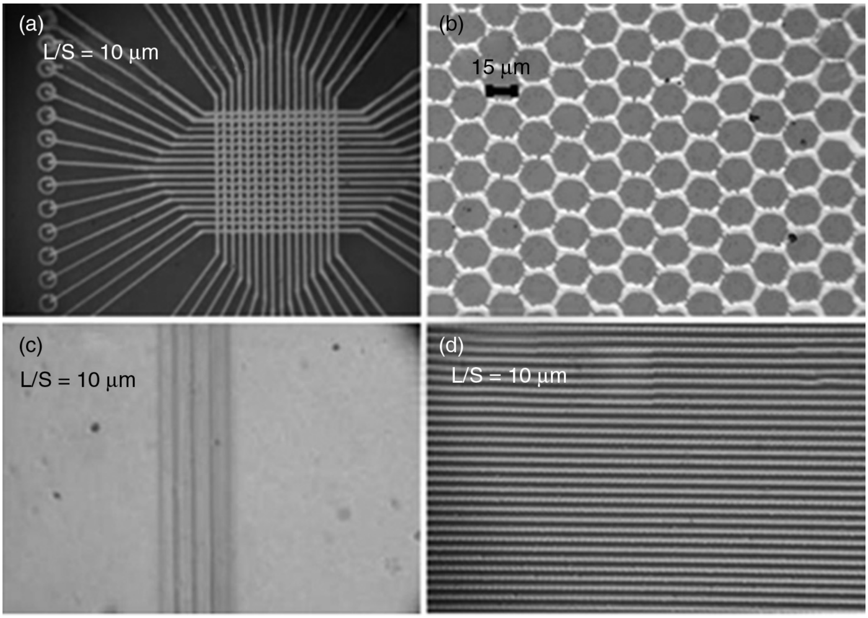
14.18 The effect on the meniscus by electrical interference. (*Source*: Si *et al.* (2007), Copyright© 2007, Elsevier.)


(a)
Control electrode
Nozzle
Insulating layer
Insulating layer
(b)
Nozzle
Insulating layer
Insulating layer

14.19 EHD multi-head printing system. (a) Gate(control) electrode EHD multi-head printing system and (b) direct EHD multi-head printing system.

14.4 Case study: EHD printing applications

The EHD printing system provides very fine, even sub-micron-scale printing and jetting, while allowing the adoption of various ink materials – such as metal, organic and biomaterials with a wider ink viscosity range of thousands of cPs. The advantages of EHD printing, further supported by the flexibility and cost-effectiveness of its manufacture, lend themselves to various



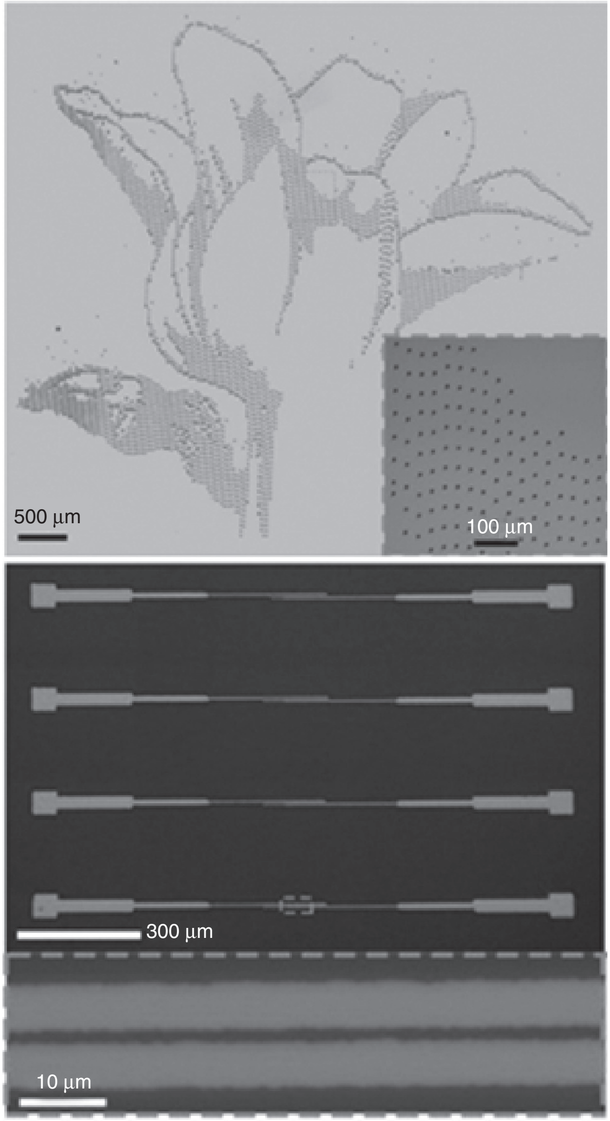
14.20 (a–d) The microscope images of fine patterns using silver NanoPasteTM. (*Source*: Murata *et al.* (2005), Copyright© 2005, Springer-Verlag.)

applications that will be sought in the future – such as the formation and/or repair of fine patterns for flat panel displays, PCBs, flexible electronics and micro-bio and -sensor chips. This section introduces the results of application research into the use of the EHD printing system.

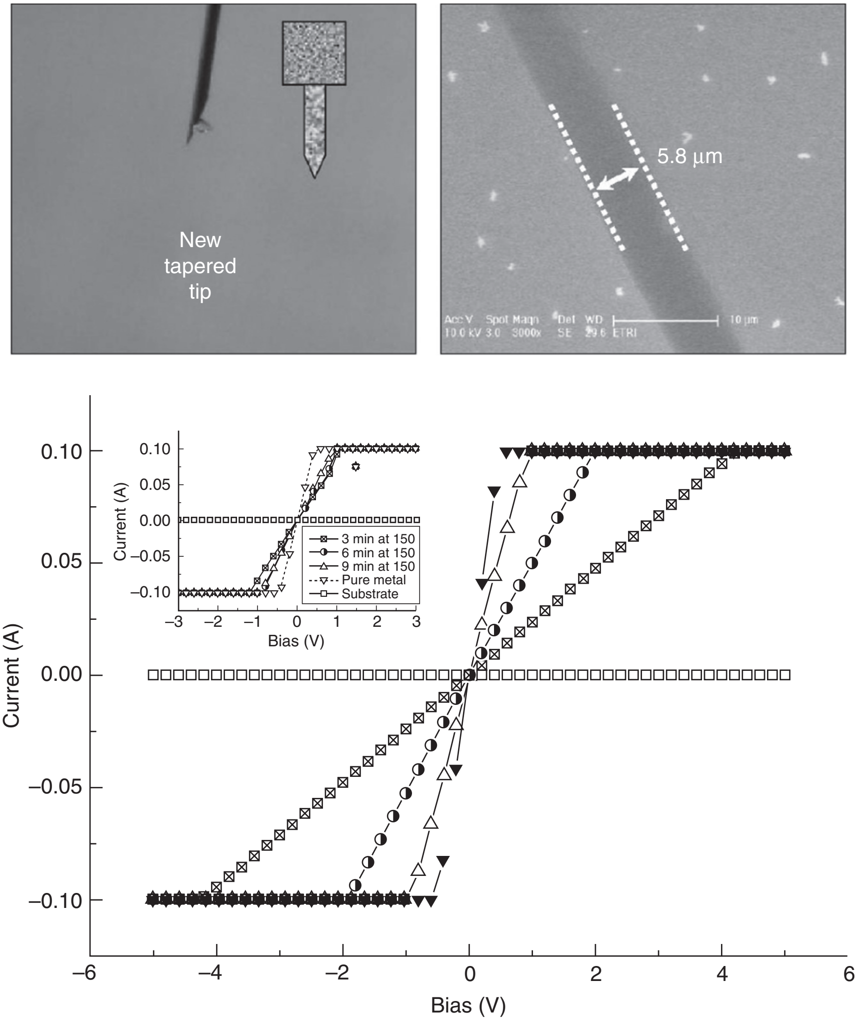
Figure 14.20 shows the circuit pattern with a line width of 3 *μ*m on an untreated substrate using the EHD printing system known as 'super-fine inkjet printing' (Murata *et al*., 2005). Figure 14.20a shows a circuit pattern which has a line width of 3 *μ*m and with 10 *μ*m pitches at the lattice area; Fig. 14.20b depicts fine wires. Figures 14.20c and 14.20d present a line width of about 3.6 *μ*m with a line space of 1.4 *μ*m and a line width of about 10 *μ*m with 20 *μ*m pitches, respectively (lines: bright; spaces: dark).

John A. Rogers at the University of Illinois at Urbana-Champaign creates complex graphic arts and high-resolution line patterns that can be applied to TFTs and flexible panel displays (FPDs) using EHD jet printing based on the EHD printing system, as shown in Fig. 14.21. This result confirms that the EHD printing system can potentially be applied in various industrial areas (Park *et al*., 2007).

Figure 14.22 illustrates line patterning approximately 6 *μ*m wide produced by the EHD printing system using a tilted-outlet nozzle, and a line pattern with a resistivity of 7×10^{-6} Ω•cm after applying an annealing process. The



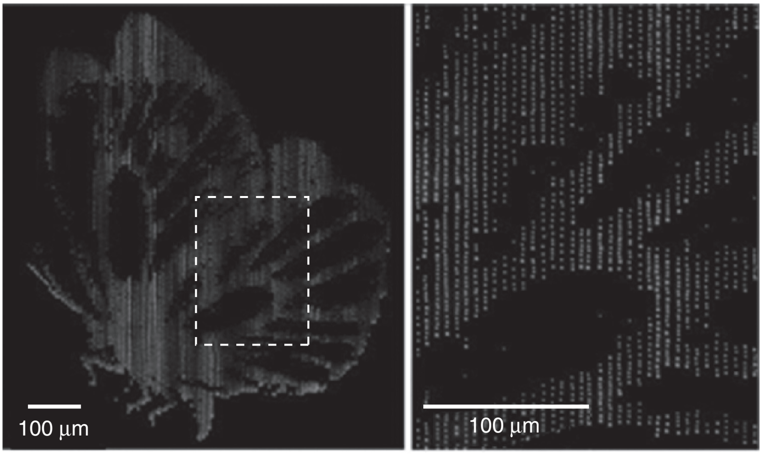
14.21 An image and an array achieved by EHD jet printing. The image of a flower was formed with printed dots of a diameter of ~8 μm diameters; the array is that of source/drain electrode pairs. (*Source*: Park *et al*. (2007), Copyright© 2007, Rights Managed by Nature Publishing Group.)



14.22 The microscopic nozzle structure of a tilted-outlet nozzle, microscopic image by a tilted-outlet nozzle (approximately 6 μm) and *I–V* curves. (*Source*: Youn *et al.* (2009), Copyright© 2009, Springer-Verlag.)

I–V curves are obtained by the 15 wt% Ag content, but the inset figure is obtained by 30 wt% Ag content (Youn *et al.*, 2009).

The EHD printing system is attracting considerable attention in the bio-research field, which requires high-resolution patterns (5–20 μm), because conventional printing systems with thermal-bubble and piezoelectric print heads cannot form high-resolution patterns. Figure 14.23 shows the fluorescence micrographs of a butterfly pattern using a nozzle with a 2 μm i.d. and



14.23 The images of patterns of DNA formed by e-jet printing based on EHD printing system. (*Source*: Park *et al.* (2008), Copyright© 2008, American Chemical Society.)

an oligonucleotide suspension (37-mer ssDNA labelled with Alexa546). This pattern is composed of arrays of dots 2 μm in diameter (Park *et al.*, 2008).

14.5 Conclusion

Industrial applications of printing technology have recently attracted considerable attention as they offer a solution to simple and low-cost fabrication and advanced pattering processes in the manufacture of electronic devices. EHD printing technology, one of the printing technologies based on EHD theory, can offer high-resolution patterning and ejection of highly viscous ink to achieve thick patterns, unlike conventional inkjet printing systems – which are thermal-bubble and piezoelectric actuators. Therefore, an EHD print head based on the direct manipulation of the liquid by an electrical field appears more promising. The EHD spray phenomenon has different ejection modes depending on the initial electrical potential and ink properties. Among them, the micro-dripping, pulsed cone-jet and continuous cone-jet modes are suitable for the EHD print head. Thus, in order to control pattern resolution and thickness, the parameters for controlling of the pattern resolution and thickness of the EHD printing system are the nozzle/substrate (electrode) geometry, electrostatic field strength, ink properties, properties of the substrate surface, patterning repetition and motor velocity.

Investigation was undertaken into the movements of the meniscus set up for the repeated ejection process to improve the structure and surface of the

EHD printing nozzle. Also, the effects of various pulse voltages, frequencies, duty cycle, and pressures for meniscus deformation and ejection were explored to find the optimal region in EHD DOD printing for ensuring repeatability and stability. The research undertaken into the EHD printing system has shown it to be an attractive candidate for fine-patterned industrial applications.

14.6 References

Allen R. R., Meyer J. D. and Knight W. R. (1985), 'Thermodynamics and hydrodynamics of thermal ink jets', *Hewlett-Packard Journal*, **36**, 21–7.

Bharathan J. and Yang Y. (1998), 'Polymer electroluminescent devices processed by inkjet printing: I. Polymer light-emitting logo', *Applied Physics Letters*, **72**, 2660–2.

Brünahl J. and Grishin A. M. (2002), 'Piezoelectric shear mode drop-on-demand inkjet actuator', *Sensors and Actuators A*, **101**, 371–82.

Byun D., Lee Y., Tran S. B. Q., Nguyen V. D., Kim S., Park B., Lee S., Inamdar N. and Bau H. B. (2008), 'Electrospray on superhydrophobic nozzles treated with argon and oxygen plasma', *Applied Physics Letters,* **92**, 093507.

Calvert P. (2001), 'Inkjet printing for materials and devices', *Chemistry of Materials.*, **13**, 3299–305.

Chen P., Chen W. and Chang S.-H. (1997), 'Bubble growth and ink ejection process of a thermal ink jet printhead', *International Journal of Mechanical Science*, **39**, 683–95.

Choi J., Kim Y., Lee S., Son S. U., Ko H. S., Nguyen V. D. and Byun D. (2008), 'Drop-on-demand printing of conductive ink by electrostatic field induced inkjet head', *Applied Physics Letters,* **93**, 193508.

Choi J., Kim Y., Son S., An K. and Lee S. (2010), 'The investigation of electrostatic induced inkjet printing system for the ejection of a stable micro/nano droplet', *NSTI-Nanotech 2010*, Anaheim, USA, 21–24 June 2010.

Choi J., Kim Y., Son S., Kim Y. M., Lee S., Byun D. Y. and Ko. H. S. (2007), 'Pattern characteristic by electrostatic drop-on-demand ink-jet printing using capillary inkjet head system', *NSTI-Nanotech 2007*, Santa Clara, USA, 20–24 May 2007.

Cloupeau, M. and Prunet-Foch, B. (1990), 'Electrostatic spraying of liquids: main functioning modes', *Journal of Electrostatics*, **25**, 165–84.

Cloupeau, M. and Prunet-Foch, B. (1994), 'Electrohydrodynamic spraying functioning modes: A critical review', *Journal of Aerosol Science*, **25**, 1021–36.

Collins R. T., Jones J. J., Harris M. T. and Basaran O. A. (2007), 'Electrohydrodynamic tip streaming and emission of charged drops from liquid cones', *Naturephysics*, **4**, 149–54.

Eggers J. (2008), 'Physics of Liquid Jets', *Reports on Progress in Physics*, **71**, 036601.

Ganan-Calvo A. M., Davila J. and Barrero A. (1997), 'Current and droplet size in the electrospraying of liquids. Scaling laws', *Journal of Aerosol Science*, **28**, 249–75.

Ishida Y., Hakiai K., Matsuzaki K., Baba A. and Asano (2005), 'Electrostatic inkjet patterning using si needle prepared by anodization', *Japanese Journal of Applied Physics*, **44**, 5786–90.

Ishida Y., Sogabe K., Haklai K., Baba A. and Asano T. (2006), 'Properties of ink-droplet formation in double-gate electrospray', *Japanese Journal of Applied Physics*, **45**, 6475–80.

Jaworek A. and Krupa A. (1999), 'Classification of the modes of EHD spraying', *Journal of Aerosol Science*, **30**, 873–93.

Jaworek A. and Sobczyk A. T. (2008), 'Electrospraying route to nanotechnology: an overview', *Journal of Electrostatics*, **66**, 197–219.

Jayasinghe S. N. and Edirisinghe M. J. (2004), 'Electrically forced jets and micro-threads of high viscosity dielectric liquids', *Aerosol Science*, **35**, 233–43.

Kim Y.-J., Choi J., Son S. U., Lee S., Nguyen X. H., Nguyen V. D., Byun D. and Ko H. S. (2010a), Comparative study on ejection phenomena of droplets from electrohydrodynamic jet by hydrophobic and hydrophilic coatings of nozzles', *Japanese Journal of Applied Physics,* **49**, 060217.

Kim Y., Ko H. S., Lee S., Son S. U., Jung D. and Byun D. (2007), 'Numerical and experimental analysis of electrostatic ejection of liquid droplets', *Journal of the Korean Physical Society*, **51**, 42–6.

Kim Y.-J., Lee S., Ko H. S., Choi J., Son S. U., An K. C. and Lee S. H., (2010b), *Discharging nozzle and electrostatic field induction ink-jet nozzle*, US patent application 0199433 A1. 2011-Aug-18.

Lastow O. and Balachandran W. (2006), 'Numerical simulation of electrohydrodynamic (EHD) atomization', *Journal of Electrostatics*, **64**, 850–9.

Le H. P. (1998), 'Progress and trends in ink-jet printing technology', *Journal of Imaging Science and Technology*, **42**, 49–62.

Lee D., Shin Y., Park S., Yu T. and Hwang J. (2007), 'Electrohydrodynamic printing of silver nanopartices by using a focused nanocolloid jet', *Applied Physics Letters,* **90**, 081905.

Lee S., Byun D., Jung D., Choi J., Kim Y., Yang J. H., Son S. U., Tran S. B. Q. and Ko H. S. (2008), 'Pole-type ground electrode in nozzle for electrostatic field induced drop-on-demand inkjet head', *Sensors and Actuators*, **141**, 506–14.

Lee S., Kim Y., Sin J., Son S., An K., Choi J., Lee W. H., Wijesundara M. and Stephanou H. E. (2011), 'Array-nozzle EHD print head and its drop-on-demand experimentation', *NSTI-Nanotech 2011*, Boston, USA, 13–16 June 2011.

Li J. L. (2007), 'EHD spraying induced by the pulsed voltage superimposed to a bias voltage', *Journal of Electrostatics*, **65**, 750–7.

Mottelay P.F. (1922), *Bibliographical History of Electricity and Magnetism,* London, Charles Griffin & Company Ltd.

Murata K., Matsumoto J., Tezuka A., Matsuba Y. and Yokoyama H. (2005), 'Super-fine ink-jet printing: toward the minimal manufacturing system', *Microsystem Technologies*, **12**, 2–7.

Notz P. K. and Basaran O. A. (1999), 'Dynamics of drop formation in an electric field', *Journal of Colloid and Interface Science*, **213**, 218–37.

Park J. U., Hardy M., Kang S. J., Barton K., Adair K., Mukhopadhyay D. K., Lee C. Y., Strano M.S., Alleyne A. G., Georgiadis J. G., Ferreira P. M. and Rogers J. A. (2007), 'High-resolution electrohydrodynamic jet printing', *Nature Materials*, **6**, 782–9.

Park J., Lee J. H., Paik U., Lu Y. and Rogers J. A. (2008), 'Nanoscale patterns of oligonucleotides formed by electrohydrodynamic jet printing with applications in biosensing and nanomaterials assembly', *Nano Letters*, **8**, 4210–16.

Si B. Q. T., Byun D. and Lee S. (2007), 'Experimental and theoretical study of a cone-jet for an electrospray microthruster considering the interference effect in an array of nozzles', *Aerosol Science*, **38**, 924–34.

Singh M., Haverinen H. M., Dhagat P. and Jabbour G. E. (2010), 'Inkjet printing–process and its applications', *Advanced Materials*, **22**, 673–85.

Sirringhaus H., Kawase T., Friend R. H., Shimoda T., Inbasekaran M., Wu W. and Woo E. P. (2000), 'High-resolution inkjet printing of all-polymer transistorcircuits', *Science*, **290**, 2123–6.

Smith D. P. H. (1986), 'The electrohydrodynamic atomization of liquids', *IEEE Transactions on Industry Application*, **IA-22**, 527–35.

Son S., An K. Kim Y. Choi J. and Lee S. (2010), 'Study of the phenomenon of meniscus deformation and ejection by pulse voltage and frequency in drop-on-demand EHD printing', *NSTI-Nanotech 2010*, Anaheim, USA, 21–24 June 2010.

Wijshoff H. (2010), 'The dynamics of the piezo inkjet printhead operation', *Physics Reports*, **491**, 77–177

Wilhelm O. (2004), Electrohydrodynamic spraying – transport, mass and heat transfer of charged droplets and their application to the deposition of thin functional films, Ph.D. dissertation, University of Tübingen.

Youn D., Kim S., Yang Y. Lim S., Kim S., Ahn S., Sim H., Ryu S., Shin D. and Yoo J. (2009), 'Electrohydrodynamic micropatterning of silver ink using near-field electrohydrodynamic jet printing with tilted-outlet nozzle', *Applied Physics A*, **96**, 933–8.

Zeleny J. (1914), 'The electrical discharge from liquid points, and a hydrostatic method of measuring the electric intensity at their surface', *Physical Review,* **3**, 69–91.

Zeleny J. (1917), 'Instability of electrified liquid surfaces', *Physical Review*, **10**, 1–6.

Zhou H. and Gué A. M. (2010), 'Simulation model and droplet ejection performance of a thermal-bubble microejector', *Sensors and Actuators B*, **145**, 311–19.

15

Photovoltaic and fuel cells in power MEMS for smart energy management

J. GARCÍA and F. J. DELGADO, University of Seville,
Spain and P. ORTEGA and S. BERMEJO,
Polytechnic University of Catalonia, Spain

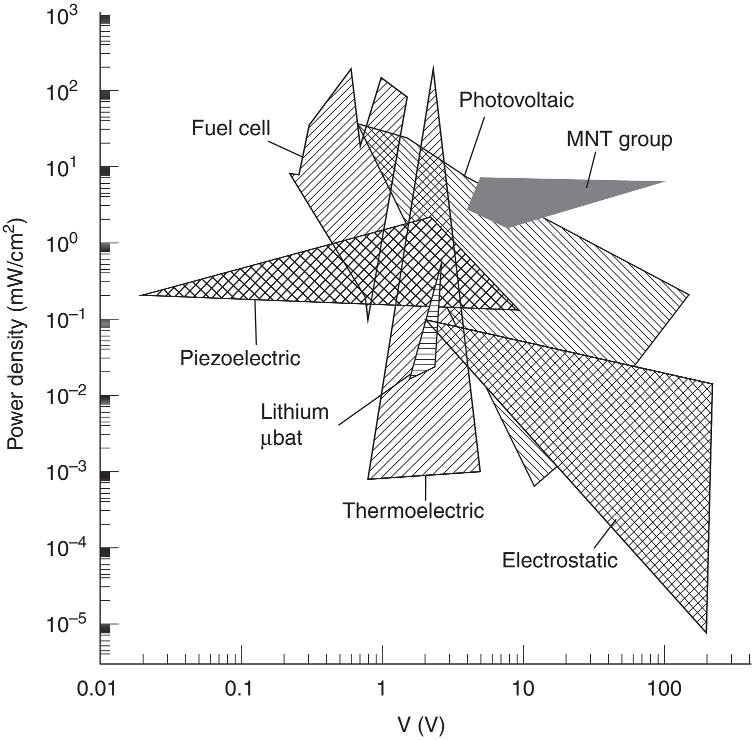
DOI: 10.1533/9780857099297.2.431

Abstract: Photovoltaic mini-generators and fuel cells are the most feasible alternatives for applications where a high density micropower supply is needed. This chapter will focus on photovoltaic and fuel-cell technologies in powering microelectromechanical systems (MEMS) for smart energy management, such as bio-implantable devices or portable electronic devices. Sun sensors are devices that are being newly introduced into photovoltaic systems in order to improve their performance. The chapter ends with a discussion of the operating principle of these sensors and their role within the photovoltaic system is described.

Key words: solar energy, photovoltaic mini-modules, energy harvesting, sun sensor, fuel cell.

15.1 Introduction

There is a growing need for small energy sources for applications such as portable electronic devices (Koeneman, 1997; Dehlinger, 2000; Liu, 2010), distributed sensors and telesupply, and in other order bio-implantable devices. Among them, MEMS supplying with a near-integrated micropower supply is a hot topic in research. All these applications demand the use of small energy sources, each one of them trying to fulfill the electrical requirements of the specific application. The different options for mini-micro power supplies found in the literature can be mainly summarized into four types of device: vibration-based (piezoelectric, electrostatic) (Sodano, 2004; Andò, 2010 Sari, 2010), thermoelectric (Muanghlua, 2000; Watkins, 2005), fuel cells (Kelley, 2006; Luque, 2010) and photovoltaic mini-generators (Lee, 1995; Shreve, 1996; Keller, 2000; Brendel, 2001; Ortega, 2001, 2008; Bermejo, 2005). Figure 15.1, which is based on some of the results shown in the review in Cook-Chennault (2008), shows a density power *vs* voltage map of all these options.



15.1 Power density and voltage map of different small-sized power supplies. The photovoltaic sources include only silicon technologies.

As can be seen, there are mainly two types of device which clearly show higher density power values: fuel cells and photovoltaic mini-generators. Of these two technologies, the photovoltaic approach has developed a wider voltage range. This can be important in applications such as the driving of MEMS switches, where relatively high voltage values are required. This chapter will focus on photovoltaic and fuel cells as power supplies for MEMS for smart energy management. However, MEMS is a quite broad term which includes a significant number of applications. This chapter will briefly describe some of these applications, starting with MEMS switches control and bio-implantable devices (Rabdill, 2010).

Sun sensors are devices that are being newly introduced into photovoltaic systems in order to improve their performance. The operating principle of these sensors and their role within the photovoltaic system are discussed at the end of the chapter.

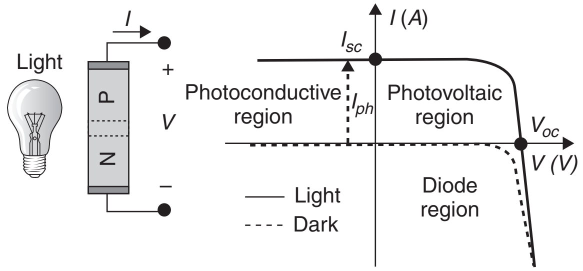
15.2 Photovoltaic mini-generators

Photovoltaic energy conversion consists in transforming light energy into electrical energy. Typically, this energy conversion is carried out in an illuminated semiconductor p-n junction diode (i.e. photovoltaic cell) in two steps. First, photons with energy greater than the semiconductor energy gap are absorbed, creating electron-hole pairs inside the device (photogeneration). Second, electron and holes pairs are separated by the electrical field existing in the p-n junction – electrons to the negative terminal and holes to the positive electrode – thus generating electrical power. Solar cells are, in fact, photovoltaic devices which are especially designed to work with solar light as an optical energy source.

In this section, the working principles of photovoltaic cells and modules will be explained. Next, several technological approaches to the fabrication of small photovoltaic modules based on crystalline silicon will be described. And finally, a number of photovoltaic applications in the fields of MEMS and autonomous systems will be explored.

15.2.1 Photovoltaic working principles

The electrical behavior of a photovoltaic cell can be studied using the superposition principle: the current–voltage *I–V* characteristics can be obtained from the corresponding characteristic of a p-n junction diode in the dark, adding a current shifting by I_{ph} (photocurrent), as shown in Fig. 15.2. Note that current leaves the device by means of the voltage reference positive terminal, the usual convention for electrical power generators.



15.2 I–V characteristics of a photovoltaic cell for dark and light conditions. The device can work in three regions depending on the actual voltage: the photovoltaic region (quadrant I), photoconductive region (quadrant II) and diode region (quadrant IV).

As shown in Fig. 15.2, photovoltaic cells can work in three possible regions depending on the actual voltage: the photovoltaic, photoconductive and diode regions:

- *Photovoltaic region* (first quadrant): this is relevant for voltages ranging from 0 to V_{oc} (open circuit voltage) – this region is where the device really can deliver electrical power to a connected load.
- *Photoconductive region* (second quadrant) for negative voltages – current is almost constant in this region and equal to the short circuit current I_{sc} corresponding approximately to I_{ph} ($I_{sc} \cong I_{ph}$); in this region, solar cells work like photodiodes being photocurrent proportional to the light irradiance and the device area. Sun sensors (described in Section 15.4) based on photodiodes (solar cells) are biased in this working region.
- *Diode region* (fourth quadrant) for $V > V_{oc}$, where the device becomes a passive component.

Photocurrent and spectral response

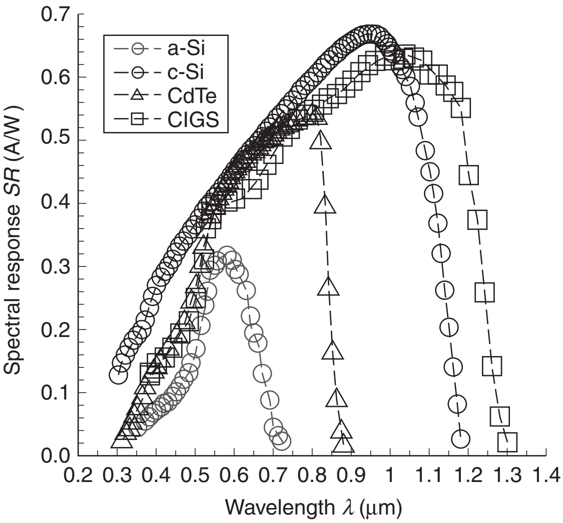
Photocurrent I_{ph} is the most important parameter in a photovoltaic cell. It can be calculated using the spectral response, denoted by *SR* in units of (A/W). Spectral response is a light wavelength λ dependent function that takes into account how much photogenerated charge leaves the contacts per second related to the incident optical power (assuming monochromatic light). Then *SR* considers any collection losses: these can be light front reflectance, bulk and surface recombination of the solar cell and partial photon absorption inside the device. These losses depend considerably on the technology and material (semiconductor) used to manufacture the device (see Fig. 15.3).

In the case of a monochromatic light of wavelength λ_o, I_{ph} can be calculated using Equation [15.1], where *S* is the optical incident irradiance – normal to cell plane, and *A* the device area.

$$I_{ph} = A \times S \times SR(\lambda_o) \qquad [15.1]$$

Alternatively, if a spectral light is used (e.g. solar light), I_{ph} is given by

$$I_{ph} = A \quad S \quad \frac{\int_0^\infty S_\lambda(\lambda) SR(\lambda) d\lambda}{\int_0^\infty S_\lambda(\lambda) d\lambda} \qquad [15.2]$$



15.3 Spectral responses *vs* wavelength for different photovoltaic semiconductor materials: monocrystalline silicon c-Si, CIGS, CdTe and amorphous silicon a-Si.

where S_λ is the spectral irradiance of the light in (W/m^2 μm) units, being the light irradiance given by:

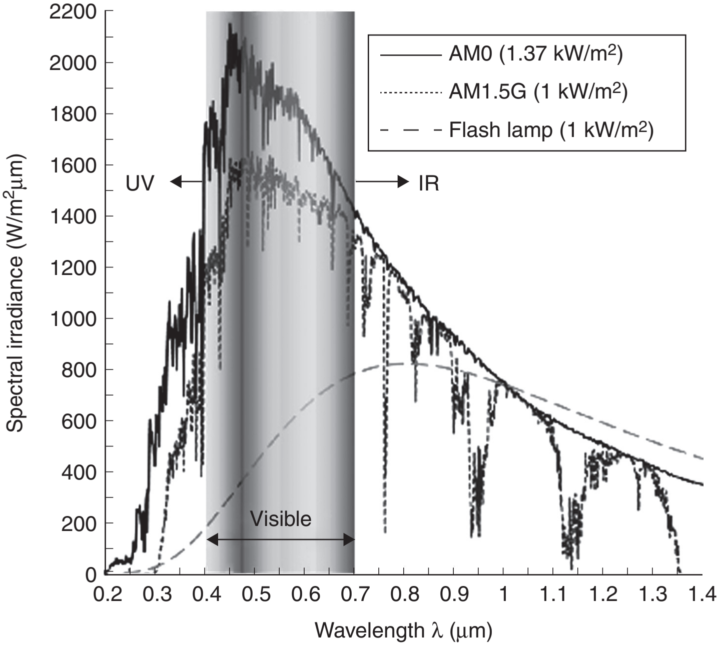
$$S = \int_0^\infty S_\lambda(\lambda) d\lambda \qquad [15.3]$$

Spectral irradiance S_λ for three-light spectrums is shown in Fig. 15.4, where the extraterrestrial AM0 (air mass) (S = 1.366 kW/m^2) and the terrestrial AM1.5G (S = 1 kW/m^2) standard solar spectrums are compared with a flash lamp light (S = 1 kW/m^2) modeled by a black body emitting source (3600 K color temperature).

I_{ph} has been calculated, for both monochromatic and spectral light, assuming normal light incidence $\gamma = 0°$ (see Fig. 15.5). In other cases, normal S irradiance must to be replaced by $S \times \cos(\gamma)$, as in Equations [15.1] and [15.2]. In order not to compromise photocurrent, and thereby decrease photovoltaic efficiency, is important to maintain γ close to 0° using, for instance, a light tracking system (as discussed in Section 15.4). This aspect is critical in systems based on light concentration.

Main photovoltaic electrical parameters

Current I and delivered power P *vs* voltage V characteristics corresponding to an illuminated cell working in the photovoltaic region are shown in



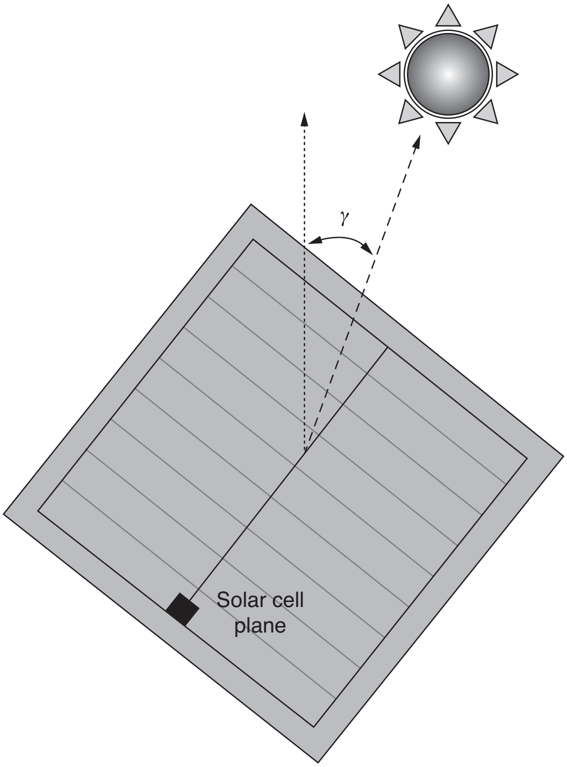
15.4 Typical light spectrums. Extraterrestrial AM0 (S = 1.366 kW/m^2) and terrestrial AM1.5G (S = 1 kW/m^2) solar spectrums. A flash lamp – assumed a black body 3600 K – (normalized to 1 kW/ m^2 irradiance) is also included for comparison. The visible light (0.4–0.7 µm range) is also depicted in the graph.

Fig. 15.6. The main photovoltaic electrical parameters are pointed out in the I–V and P–V plots. Of these, short circuit current I_{sc} and open circuit voltage V_{oc} have already been introduced. Additionally, other important parameters are the maximum delivered power P_m, and the current and the voltage at the maximum power point (mpp point), I_m and V_m, respectively.

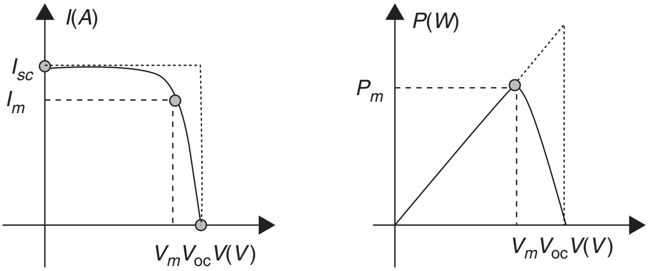
Another important parameter is the fill factor (FF) given by:

$$\mathrm{FF} = \frac{P_m}{V_{\mathrm{oc}} I_{\mathrm{SC}}} \qquad [15.4]$$

This parameter indicates how far a photovoltaic cell is from ideal electrical performance (see the ideal curves represented by dotted lines in Fig. 15.6). Finally, the photovoltaic energy conversion efficiency η can be



15.5 Definition of the light incidence angle γ.$\gamma = 0$ means normal light incidence.



15.6 Characteristics of current–voltage *I–V* and power-voltage *P–V*. *I–V* (left) and power-voltage *P-V* (right). The photovoltaic main parameters are pointed out in the graphs, as these can be V_{oc}, I_{sc}, V_m, I_m and Pm. Ideal *I–V* and *P–V* characteristics are also shown (dotted lines).

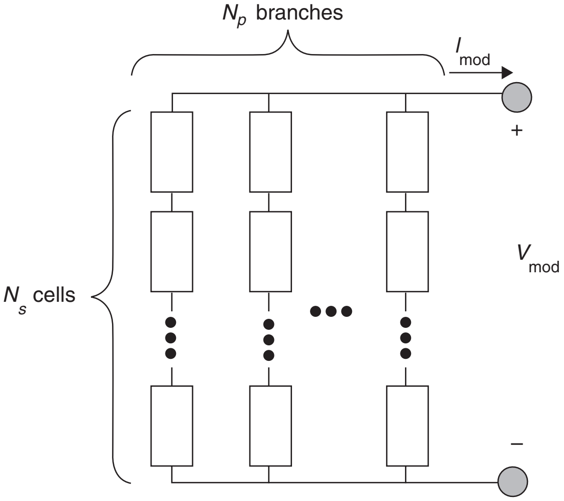
calculated taking into account incident light power (P_{in}) or, alternatively, considering irradiance S and cell area A, by:

$$\eta = \frac{P_m}{P_{\text{in}}} = \frac{\text{FF} \times V_{\text{oc}} \times I_{\text{SC}}}{S \times A} \qquad [15.5]$$

Photovoltaic modules

In order to scale up current and voltage to fulfill electrical requirements, cells are combined electrically in series to raise the voltage and in parallel to increase the current (see Fig. 15.7). Table 15.1 summarizes the scaling rules of an ideal module formed by N_p branches in parallel with N_s identical cells connected in series.

Module current can be increased by enlarging the cell area and then increasing photocurrent I_{sc} (see Equation [15.1]) or by connecting several



15.7 An ideal module consisting of a connection of N_p branches in parallel. Each branch is formed by N_s identical cells in series.

Table 15.1 Scaling up rules of an ideal module consisting of N_p branches in parallel of N_s identical cells in series

$V_{\text{oc\_mod}}$	$I_{\text{sc\_mod}}$	$V_{\text{m\_mod}}$	$I_{m\text{\_mod}}$	$P_{m\text{\_mod}}$	FF_{mod}	η_{mod}	A_{mod}
$N_s \times V_{oc}$	$N_p \times I_{sc}$	$N_s \times V_m$	$N_p \times I_m$	$N_p \times N_s \times P_m$	FF	η	$N_p \times N_s \times A$

branches in parallel (i.e. $N_p > 1$). However, scaling up voltage can only be carried out by combining cells in series ($N_s > 1$). For instance, if monocrystalline silicon material is used, a single solar cell provides approximately 0.6 V in open circuit conditions ($T = 25°\text{C}$ AM1.5G 1 kW/m^2 spectrum) or ~0.45 V at the mpp point. Thus, a typical module arrangement of 32 cells in series is required to feed a 12 V battery. As can be seen from Table 15.1, ideally, module conversion efficiency is the same as that of the single solar cell. However, this is not true of an actual module because, due to a useless module area (A_{mod}), from the perspective of photovoltaic conversion, it is necessary to perform isolation and electrical interconnections between cells (i.e. $A_{\text{mod}} > N_p \times N_s \times A$).

15.2.2 Mini-modules technologies

Compact and efficient small area photovoltaic arrays (<1 cm^2) can be fabricated using different materials and technologies: (a) gallium arsenide (GaAs) substrate (Beaumont, 1991), (b) thin-film amorphous silicon (a-Si) on a dielectric substrate (Lee, 1995) or (c) crystalline silicon (c-Si) (Ortega, 2001, 2002, 2008; Bermejo, 2005). Table 15.2 summarizes the results published so far in the literature, where N_s is the number of cells connected in series, V_{oc} is the module open circuit voltage, v_{oc} and p_m are the 'specific open circuit voltage' and 'specific maximum power' respectively, first introduced here and defined as the ratios of V_{oc} and P_m to the total module surface, respectively. These are indicators of performance per unit area.

Although GaAs mini-modules can be very efficient (both v_{oc} and p_m levels are high), they are more expensive to produce compared with mainstream silicon technology. Alternatively, a-Si arrays can be subject to long-term stability problems. Therefore, c-Si material appears to be a good choice for the fabrication of very small photovoltaic arrays.

Table 15.2 Comparison between mini-module fabrication technologies

Technology	Area (cm^2)	N_s	$V_{\text{oc\_mod}}$ (V)	v_{oc} (V/cm^2)	p_m (mW/cm^2)	References
GaAs[a]	0.01	6	6.5	650	1204	Beaumont, 2001
a-Si	1.00	100	150	150	0.21	Lee, 1995
c-Si/MCM	0.70	9	5.1	7.3	7.7	Ortega, 2002
c-Si/fusion-bonding	1.40	9	4.1	2.9	2.7	Bermejo, 2005
c-Si/SOI	0.43	169	100	240	7.3	Ortega, 2008

[a] Measured using a beam laser of wavelength $\lambda = 810$ nm and irradiance 22.80 kW/cm^2 as an optical source. Other cases are measured using standard solar spectrum AM1.5 (1 kW/cm^2).

Devices based on c-Si can be fabricated in a number of ways – for example, non-monolithic integration using microelectronic packaging techniques such as multi chip module (MCM) technology (Ortega, 2001). More interestingly, especially where integration and compatibility with other silicon devices is involved, a monolithic design (Bermejo, 2005; Ortega, 2008) reduces fabrication steps and reliability is enhanced. Generally, monolithic integration is achieved by means of total (Bermejo, 2005) or partial (Ortega, 2008) anisotropic etching of the silicon substrate, forming channels that isolate individual cells from each other. As can be seen in Table 15.2, a monolithic integration using silicon-on-insulator (SOI) technology is one of the best alternatives (using c-Si) to obtain very high specific voltages with relatively high specific powers.

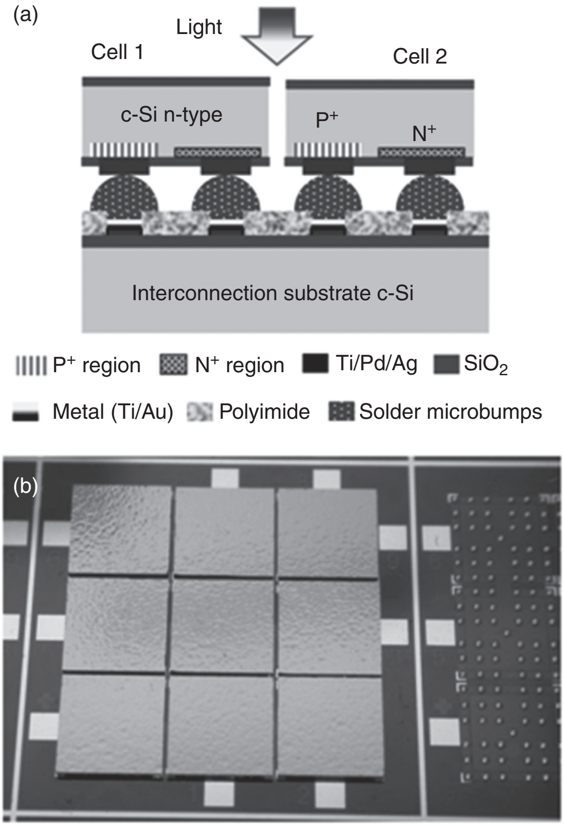
In the following subsections, the silicon-based photovoltaic mini-module technologies developed since 2000 by the Micro and Nano Technology (MNT) Group are introduced and details are given regarding certain technological and electrical results.

Multichip module (MCM) technology

This approach use c-Si solar cells, batch-produced, in combination with MCM flip-chip technology (Al-Sarawi, 1998). The MCM approach (Ortega, 2001) is based on the interconnection of c-Si back contact photovoltaic cells in series by means of a common c-Si substrate wafer, as shown in Fig. 15.8a.

MCM mini-module fabrication is divided into three phases: fabrication of the solar cells, fabrication of the substrates and flipping of the solar cells:

- *Fabrication of the solar cells*. These are made using c-Si wafers (device-wafer); single solar cells are obtained by dicing the device wafer using a diamond circular saw.
- *Fabrication of the substrates*. These are used for the interconnection between solar cells and are made from c-Si wafers (substrate wafer).
- *Flipping of the solar cells*. The solar cells are flipped, picked and placed on top of a substrate to be assembled with an alignment accuracy of 5 μm. A separation between dice (solar cells) in the substrate of around 100 μm is necessary in order to perform the pick and place stage adequately. Before placing the chips into the substrate, a solder paste is deposited onto the substrate contacts using screen printing technology. Many different solder pastes are available in the market. They can be lead-based like Sn/Pb and Sn/Pb/Ag, and lead-free like Sn/Ag, Sn/Sb, In, Pd, Sn/Ag/Cu. The metallic paste is converted into a microbump (~50 μm high) after a reflow step at a temperature well above the melting point, ranging from 180°C to 320°C depending on the selected alloy. Once the solder paste has melted, the chips are attached to the


(a)
Light
Cell 1
Cell 2
c-Si n-type
P+
N+
Interconnection substrate c-Si
P+ region
N+ region
Ti/Pd/Ag
SiO2
Metal (Ti/Au)
Polyimide
Solder microbumps
(b)

15.8 (a) A sketch of the MCM mini-module concept showing the interconnection of two adjacent solar cells and (b) a fabricated module, before dicing, of nine solar cells electrically connected in series with a total area of 2.8 mm × 2.8 mm. On the right-hand side of the image is a substrate where no solar cells were placed, as can be seen by the solder microbumps.

substrates. Finally, the substrate wafer has to be diced in order to obtain the MCM mini-modules.

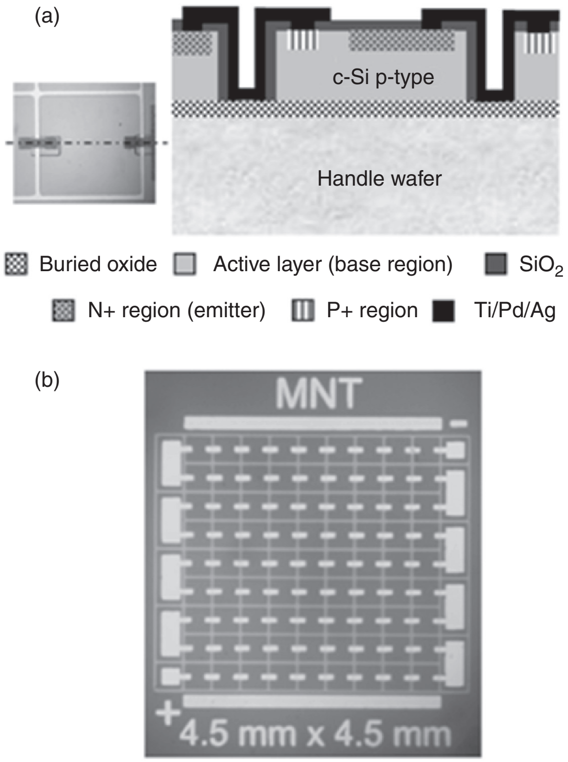
Flip-chip technology requires that the cell contacts (emitter and base contacts) are placed in the same surface (rear surface) so that the device and substrate contacts are face-to-face. As a consequence, light must enter into the device by the opposite surface to the contacts (the front). Therefore, photogenerated minority carriers have to travel the thickness of the wafer to arrive at the p-n junction at the rear of the unit. In this type of solar cell, known as a back contact solar cell (Van Kerschaver, 2006), it is important to have a very well passivated front surface (i.e. low recombination surface velocities) and high bulk lifetimes to guarantee a high carrier collection (i.e. high photocurrents).

Figure 15.8b depicts a fabricated MCM square module of nine solar cells connected in series with an area of 2.8 mm × 2.8 mm, obtaining an open circuit voltage of almost 5.1 V with a conversion efficiency of 10.6% under a standard solar spectrum (AM1.5 1 kW/m^2). One of the main drawbacks of using this technology to fabricate small photovoltaic modules is the minimum solar cell size necessary to perform the pick and place stage, which is restricted to ~1 mm^2. However, in applications where very small mini-modules are not required, merging silicon solar cells with mass production flip-chip assembly can significantly reduce the mini-module fabrication cost.

Silicon-on-insulator (SOI) technology

In this approach, solar cells and mini-modules are fabricated monolithically together using off-the-shelf SOI c-Si wafers. As Fig. 15.9a shows, photovoltaic cells are processed on an active layer of a relatively small thickness (3–50 μm) on top of a handle wafer, which acts as a mechanical support; in between the active layer and the handle wafer there is a thick SiO_2 buried oxide layer (BOX), typically 1 μm thick. The top active layer is where individual solar cells are fabricated (see Ortega (2008) for details). An electrical isolation between devices is mandatory; this is carried out by means of channels created by anisotropic etching of the whole active layer. Both the emitter and base contacts are on the front of the wafer, making series interconnection between cells easy; this is performed by metal evaporation and patterning. A thin thermal oxide is grown on the front surface of the device to passivate and to obtain an antireflective coating film. An important difference with respect to MCM technology is that light enters into the device on the same face where the base and emitter contacts are placed. Thus, the photogeneration mechanism occurs near the p-n junction, avoiding carrier recombination losses. However, metallization of the front should be reduced to a minimum to avoid shadow light losses.

SOI wafers present a design challenge: in order to guarantee an acceptable photogeneration under either monochromatic or spectral light, the active layer should be designed to be sufficiently thick. On the other hand, thick active layers require deep channels, thereby causing possible interconnection problems; there is therefore a trade-off between the thickness of the active layer and the fabrication yield. This technology is very well-suited to UV and visible light (monochromatic or spectral) due to the high absorption coefficient of silicon in the 350–700 nm wavelength range; in this instance, photons are absorbed very close to the front surface. Figure 15.9b shows a fabricated mini-module consisting of 81 cells, each one being 0.225 mm^2, and the total module area being 0.45 × 0.45 cm^2. In this mini-module, an

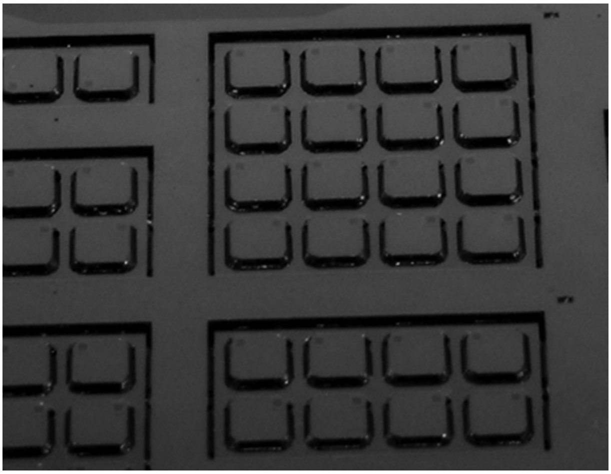


15.9 (a) Top view of an SOI solar cell (left) indicating the cross-section that was sketched to show the concept of the SOI mini-module (right) and (b) a fabricated module of 81 solar cells monolithically connected in series with a total area of 4.5 mm × 4.5 mm.

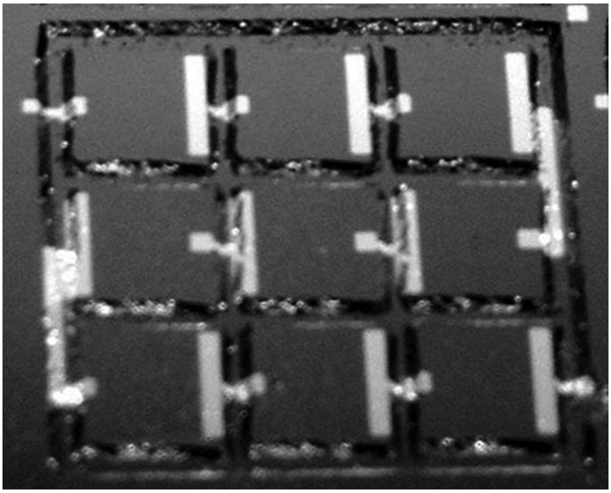
open circuit voltage of 57 V has been measured with a conversion efficiency of 9.7% using standard solar spectrum AM1.5 1 kW/m^2. It is worth noting that, in this mini-module fabrication technology approach, very high specific open circuit voltages (~240 V/cm^2) can be achieved.

Fusion-bonding technology

This concept is similar to the SOI technology explained in the previous section, though in this case the active layer is 350 μm thick. Like in the SOI approach, two wafers are used: a silicon wafer (also known as the device wafer), where the solar cells are fabricated, and the handle wafer. The device wafer is bonded or stuck to the handle wafer using fusion-bonding or adhesive bonding techniques. Fusion-bonding has been chosen as an appropriate alternative, using a second silicon wafer as a substrate material. The handle wafer provides electrical isolation and mechanical stability to the devices placed in the silicon wafer on the top. Electrical isolation between solar cells is essential because they share the same substrate, so individual cells are isolated by channels created by anisotropic etching of the silicon, and series are connected to the neighboring cell by means of the metal pattern. The schematic view is similar to that shown in Fig. 15.9a.



15.10 Closed view of a 2″ processed wafer with different cell topologies before deposition of the interconnection path.



15.11 Mini-module of nine cells in series processed with fusion-bonding technology. The total area is 1.4 cm^2.

Monolithic mini-module fabrication has three main steps: the bonding process of the device and the substrate wafer, the fabrication of the isolation channels and the serial connection between individual cells by metal tracks. Figure 15.10 presents a closed view of a 2″ processed wafer exhibiting different topologies before the deposition of the metallization tracks. Figure 15.11 shows a photovoltaic mini-module with nine cells connected in series. The total area is 1.4 cm^2; the total open circuit voltage is 4.1 V and the specific power density is 2.7 mW/cm^2.

The main advantage of this technology compared with the SOI is the increase in light absorption, thus increasing the available current density. The main drawback of this technology is the challenging fabrication procedure, where critical points such as the photolithography and metal track deposition over 350 μm grooves may need to be resolved.

15.3 Applications of photovoltaic mini-generators

There are many applications which demand autonomous power supplying. Wireless telesupplying is mandatory in harsh environments or in places where electrical conduction is not possible, such as in satellites or in bio-implantable systems. Consumer electronics will also benefit as well from a close, suitable mini-energy source. MEMS switches demand a relatively high voltage and low current values, which make the use of a specific mini-energy source an interesting choice. In this chapter, we show two example applications: (a) telesupplying bio-implantable devices and (b) MEMS driving. In these applications, the introduction of a small-sized photovoltaic mini-module significantly modifies the energy balance of the system, whether energy harvesting solar light or using monochromatic light.

15.3.1 Telesupplying bio-implantable devices

There are many applications where the use of a complementary small-sized energy source that is capable of wirelessly giving an extra energy contribution is of interest. Bio-implantable devices like pacemakers (Dinesh, 2010), implantable defibrillators (Andrew, 2010), drug deliverer systems (Nisar, 2008) or neurotransmitters (Yoshimi, 2004), for example, would benefit from the extra energy supplied by the photovoltaic mini-module, thus increasing the useful life of the implantable battery.

The design of a suitable energy source goes through the definition of the specific voltage and current requirements of the whole system. In an autonomous device, the initial capacity of the battery can be improved by means of an external energy source. Calculating the energy balance of the total system, the total capacity of the autonomous device is given by Equation [15.6].

$$C(t)\big|_{t=M}[Wh] = C(0)[Wh] + (P_{\mathrm{PV}} - P_L)[Wh/\mathrm{year}]M[\mathrm{year}] \qquad [15.6]$$

where $C(0)$ is the initial battery capacity, P_L is the power consumption, P_{PV} is the power given by the photovoltaic mini-module and M is the number of autonomy years. The power produced by a photovoltaic source is defined as:

$$P_{\mathrm{PV}}[Wh] = \int_0^t S[W/\mathrm{cm}^2]A_{\mathrm{mod}}[\mathrm{cm}^2]\,\eta dt \qquad [15.7]$$

where S is the irradiance, A_{mod} is the photovoltaic mini-module surface and η the photovoltaic efficiency. Taking S, A_{mod} y η constant in time, the stored energy as a function of the exposition time, H_{exp}, is:

$$P_{PV}[Wh/\mathrm{year}] = S[W/\mathrm{cm}^2]A_{\mathrm{mod}}[\mathrm{cm}^2]\,\eta H_{\mathrm{exp}}[h/\mathrm{week}]\frac{1\,\mathrm{week}}{7\,\mathrm{days}}\frac{365\,\mathrm{days}}{1\,\mathrm{year}} \qquad [15.8]$$

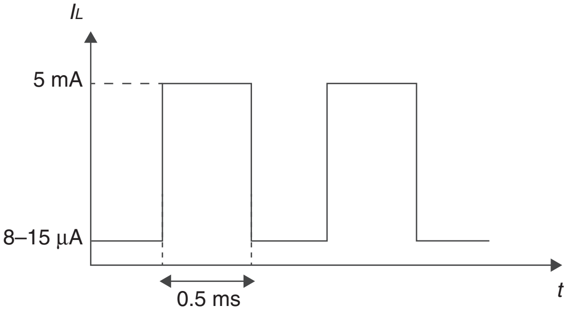
Substituting Equation [15.8] in [15.6] and imposing the final capacity as being null, the surface can be evaluated by:

$$A_{\mathrm{mod}} = \left(P_L - \frac{C(0)}{M}\right)\frac{7}{S\eta H_{\mathrm{exp}}365} \qquad [15.9]$$

where M can be expressed as a function of the increase of battery life P and the nominal duration of the battery without any provision of external energy, following the expression:

$$M[\mathrm{years}] = N[\mathrm{years}]\left(1 + \frac{P[\%]}{100}\right) \qquad [15.10]$$

Figure 15.12 shows the real consumption of a commercial pacemaker with a Wilson Greatbatch WG 8402 lithium iodine battery ($C(0)$ = 2.8 Wh, V_L = 2.8 V, f = 60–70 Hz). Table 15.3 shows the photovoltaic area necessary to fulfill the consumption of the pacemaker for different illumination conditions,



15.12 Current consumption of a commercial pacemaker with a Wilson Greatbatch WG 8402 lithium iodine battery.

Table 15.3 Photovoltaic area necessary to increase the battery lifetime up to *P* of the pacemaker consumption shown in Fig. 15.12

Light source	η (%)	*S* (kW/m^2)	Exposition time (h/week)	*P* (%)	Area (cm^2)
Solar AM 1.5	4.3	1	1	60	1
Monochromatic λ = 830 nm	10.5	0.18	5	60	1.2
Low consumption bulb OSRAM 23 W (distance 5 cm)	5	0.115	2.5	20	1

photovoltaic efficiencies, irradiance and exposition time in order to increase the battery lifetime to *P*. For instance, if we consider a standard AM1.5 solar spectrum (1 kW/cm^2) and 4.3% of photovoltaic efficiency, a 1 cm^2 solar cell can increase the battery life by up to 60%, using an exposition time of only 1 h per week.

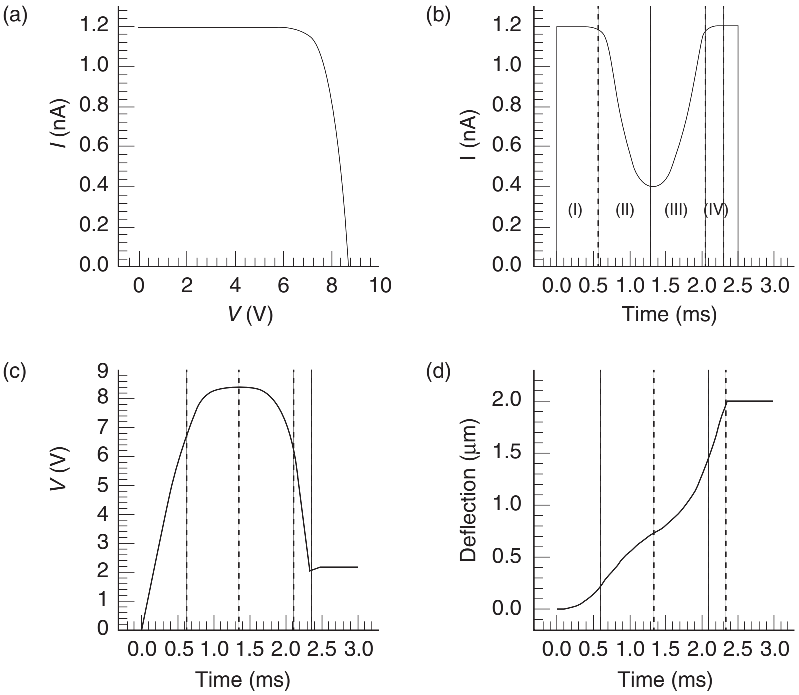
15.3.2 Driving MEMS switches

MEMS electrostatic actuators are currently used for a number of applications such as optical switching, electrical relays or radio frequency (RF) variable capacitors for example. Driving such devices is usually performed by applying a voltage which produces an electrostatic force which is nonlinearly related to the deflection of the moveable part of the actuator and usually requires higher voltage than conventional standard processing electronics. The nonlinearity involved makes movement unstable, leading to the collapse of a device at a so-called 'pull-in' voltage (Nathanson, 1967). It has been observed (Castañer and Senturia, 1999) and experimentally demonstrated (Castañer *et al.*, 1999) that a 'charge drive', instead of a voltage drive, has a number of benefits, as the total power consumed in the switching can be drastically reduced, the kinetic energy at the impact of the two plates is reduced by orders of magnitude and the movement can be analog-controlled across the gap (Castañer *et al.*, 2001). It has also been shown that a charge drive usually requires high voltage compliance (Seeger, 1997).

Actually, a photovoltaic mini-array can be considered as a new DC/DC step-up converter without the need of conventional charge pumps of switching converters, avoiding the use of switched-capacitor high-frequency circuits or bulk inductors. The series connection of solar cells allows the scaling up voltage in the array to levels more suitable for MEMS actuation (typically more than 5 V). In this section, we will to briefly show the potential use of a photovoltaic source as the direct drive of electrostatic actuators. Since the *I–V* characteristic of a photovoltaic array has the shape of a current source limited in voltage, the actuation process can be that of a charge drive, voltage-limited. The main potential benefit we foresee is that a direct drive with galvanic isolation can be achieved at high voltages without using any typical DC/DC converter. This photovoltaic drive therefore becomes a new inductor-less DC/DC type of converter 'opto-activated'. The number of individual solar cells can be chosen to drive specific MEMS switches of high pull-in voltage value.

A PSPICE model has been developed (Bermejo, 2005) including the photovoltaic mini-module and an electrostatic switch. Figure 15.13 shows the main parameters of the transient. The transient starts in t = 0s in Fig. 15.13. As soon as the irradiance pulse is turned ON, the MEMS device capacitance starts charging with a current of value I_{sc} and the moveable arm starts moving due to the increase of charge and, hence, of the attractive force. For a while, the current remains reasonably constant while the voltage rises (Region I in Fig. 15.13), and the operating point will move from left to right following the *I–V* characteristics (Fig. 15.13a). When the operating point goes beyond the flat part of the stationary characteristic in Fig. 15.13a, the voltage (Fig. 15.13c) will still be increasing, but more slowly, and the current (Fig. 15.13b) will start dropping (Region II). At some point, the charge in the MEMS device reaches the value of the pull-in charge, which is required for the actuator to collapse. At this moment, the moveable arm moves drastically producing a large change in the value of the actuator capacitance. As the current has a limiting value, the voltage across the MEMS device has to drop (this is seen in Regions III and IV), as the total charge cannot increase further. This drop in the voltage forces an increase in the current and the operating point now changes direction and moves backwards over the *I–V* curve finally reaching a hold-on voltage considerable smaller than the maximum drive-in voltage.

Analyzing the energy at the impact, a decrease of up to three orders of magnitude, for the same switching time, can be obtained by using this current-controlled method.



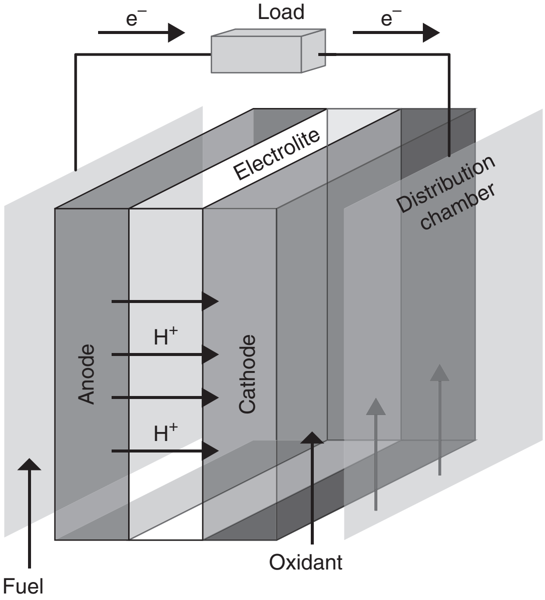
15.13 PSpice simulation results of the photovoltaic source and the MEMS switch models. (a) *I–V* of the photovoltaic source, (b) current delivered to the switch, (c) switch voltage and (d) deflection of the movable arm.

15.4 Micro-fuel cells

A fuel cell is a device that converts chemical energy into electrical energy. Like in photovoltaic cells, this energy conversion is direct, so no thermodynamic cycle is involved. Unlike a battery, a fuel cell produces electric energy as long as fuel is supplied and has a higher density of stored energy. Both properties make the fuel cell a very interesting power source for portable equipment.

15.4.1 Fuel-cell principles and classification

A fuel cell has a very simple electromechanical structure, so it is easy to implement using MEMS technology. The basic structure of a fuel cell consists of two electrodes (cathode and anode) connected through an electrolyte. The oxidant and the fuel are combined internally to produce electrons, which are driven through an external electric load. A distribution chamber



15.14 Basic structure of a fuel cell.

provides the fuel and the oxidant to the electrodes and drives electrons through an external load (see Fig. 15.14).

An electrolyte is a chemical medium with high ionic conductivity but low electronic conductivity. The most widely used electrolyte is a membrane of poly-tetrafluoroethylene (PTFE): Nafion®. This membrane has good mechanical and chemical properties such as high mechanical strength, ease of handling and high stability. Other usual electrolytes are aqueous alkaline solution (AKS), phosphoric acid (PO_4H_3), lithium potassium carbonate ($C0_3$ Li K) or yttria-stabilized zirconia (YSZ).

Electrodes are in direct contact with both surfaces of the membrane. In the anode, fuels like hydrogen are dissociated in ionic species (in this case, protons) and electrons.

$$H_2 \rightarrow 2H^+ + 2e^- \quad [15.11]$$

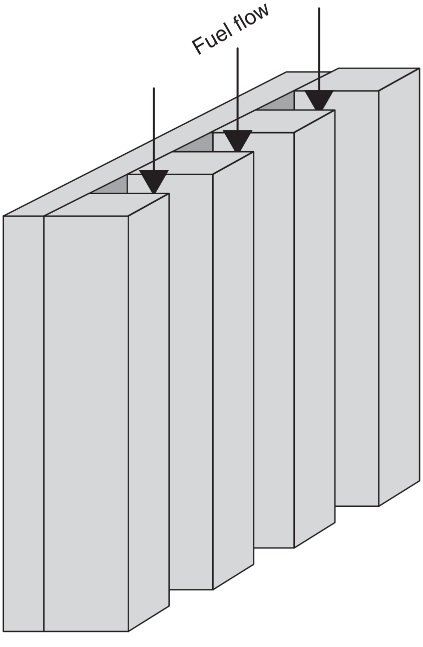
Protons are driven to the cathode through the membrane, while electrons travel to the cathode through the external load. In the cathode, the oxidant (typically oxygen), protons and electrons react to produce water, in this case.

$$2H^{+} + 2e^{-} + \frac{1}{2}O_2 \rightarrow H_2O \qquad [15.12]$$

The desirable properties of electrolytes include their high porosity and the catalyst power. Porosity gives the protons higher mobility and the catalyst power shifts the reaction to the product side. Electrodes are usually composed of a mixture of platinum, carbon and Nafion®.

The last components of a fuel-cell structure are the distribution chambers. They constitute the external shield of the system and are in charge of providing fuel and oxidant to electrolytes, the removal of the product of reaction (water) and connecting the cell electrically to external devices. The distribution chamber requires high electrical and thermal conductivities, mechanical strength and corrosion resistance. Graphite is a good material because it has good electrical and thermal conductivities, but unfortunately is too fragile. In order to drive the fuel and the oxidant efficiently, electrodes have to be channeled. Those channels are also used to remove the resulting water from cathode; this is an important target in the effort to improve fuel cells (see Fig. 15.15).

The most common fuel is hydrogen but several hydrocarbons are used too, such as methanol or ethanol. Despite the wide variety of fuels, the



15.15 Channeled structure of an electrode.

Table 15.4 Properties of the fuel cells most usually employed

Fuel-cell type	Electrolyte	Electrolyte state	Charge carrier	Operating temp (°C)	Power range (kW)
(PEM) polymer electrolyte membrane	Sulfonated PTFE	Solid	H^+	50–80	5–250
(AFC) alkaline	AKS	Liquid	OH^-	90–100	5–150
(DM) direct methanol	H^+ interchange membrane	Solid	H^+	80	5
(PAFC) phosphoric acid	PO_4H_3	Liquid	H^+	160–220	$50–1 \times 10^4$
(MCFC) molten carbonate	$C0_3$ Li K	Liquid	CO_3^{2-}	620–660	$100–2 \times 10^3$
(SOFC) solid oxide	YSZ	Solid	O^{2-}	800–1000	100–250

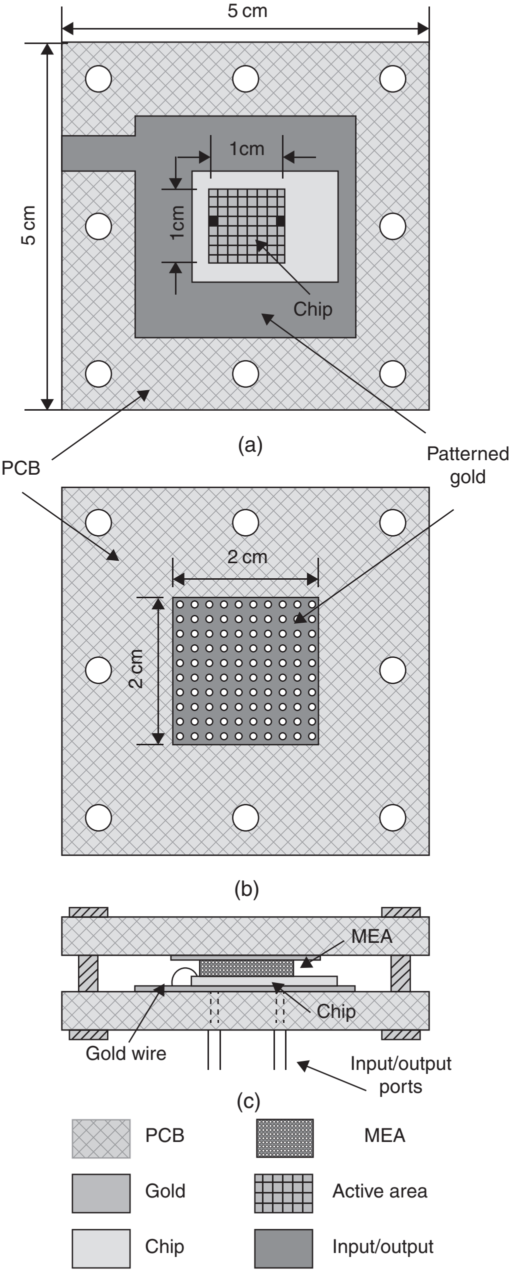
classification of fuel cells is based on the type of electrolyte used in their structure. Table 15.4 shows the most widely used fuel cells, their properties and their operational conditions.

The voltage of a single fuel cell is approximately 0.7 V in normal operating conditions, so several cells have to be stacked in series to increase the output voltage or in parallel to produce more electrical current. The efficiency of a fuel cell typically ranges from 40% to 60% depending on how efficient several processes are; for instance, waste water removal, humidity of the membrane, heat evacuation and fuel supply. All these factors must be taken into consideration in the design and manufacturing technology of a fuel cell.

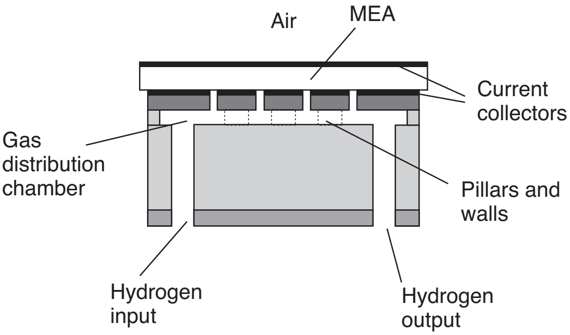
15.4.2 MEMS-based polymer electrolyte membrane (PEM) fuel cell

At the time of writing the trend is to miniaturize power sources. The use of MEMS technology for fuel-cell design allows decreased dimensions and reduced losses, thereby improving energy conversion and fuel storage efficiency. Many efforts have focused on developing a MEMS-based process in order to miniaturize fuel-cell structure.

Figure 15.16 shows the structure of a polymer electrolyte membrane (PEM) fuel cell in which an integrated silicon distribution chamber has been developed (Luque, 2010). The use of silicon technology has several advantages because it uses well-known fabrication processes. Silicon is a conductive material which will collect generated current from the anode and the resulting monolithic structure could be integrated with auxiliary electronics.


5 cm
5 cm
1cm
1cm
Chip
(a)
PCB
Patterned gold
2 cm
2 cm
(b)
MEA
Chip
Gold wire
Input/output ports
(c)
PCB
MEA
Gold
Active area
Chip
Input/output

15.16 PEM fuel-cell design. (a) Overhead view of the PCB bottom cover with the fuel-cell chip, (b) overhead view of the top cover with the current collector and (c) cross-sectional view of the complete assembly.



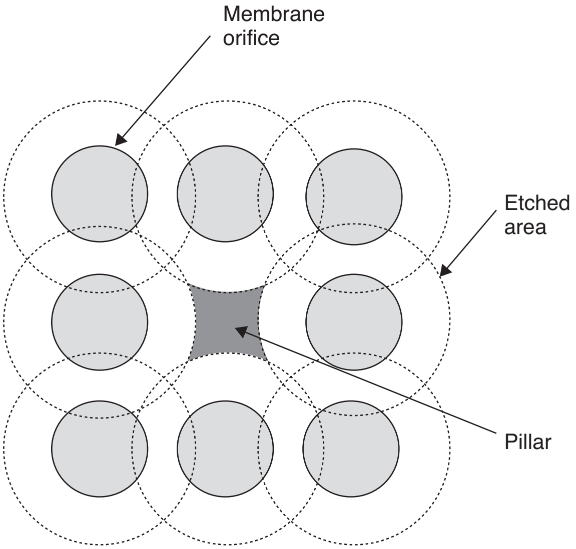
15.17 General structure of a silicon fuel cell with the integrated fuel distribution chamber.

Figure 15.17 shows the structure of the fuel distribution chamber. Two channels have been implemented to supply hydrogen to the membrane and to evacuate the waste hydrogen not consumed. The distribution chamber has many holes to connect it with the membrane electrode assembly (MEA). A catalyst, made of platinum, is in contact with the MEA and promotes hydrogen dissociation. The generated electrons are collected by a silicon structure and protons travel through it to reach the 'MEA'.

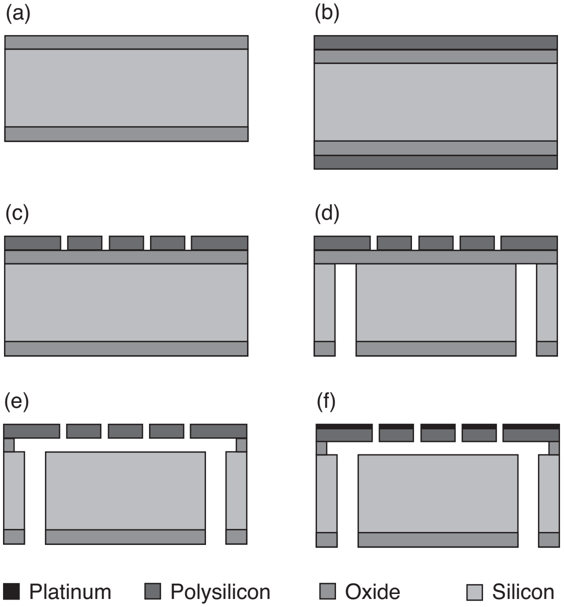
Two types of structures have been implemented in the distribution chamber: pillars and walls. Pillars were built to support the membrane and prevent it from failing when hydrogen flows below it. Walls are used to force the hydrogen to follow a particular path when flowing and thus force it to cover a greater area of the membrane. Both were made by patterning the sacrificial oxide of silicon and a particular distribution of holes, as shown in Fig. 15.18.

The fabrication process, based on a previous work (Kelley, 2002), is described by Luque (2010). Figure 15.19 shows how the process starts with a p-doped double-sided polished Si wafer. The steps in the fabrication process are summarized in points a–f:

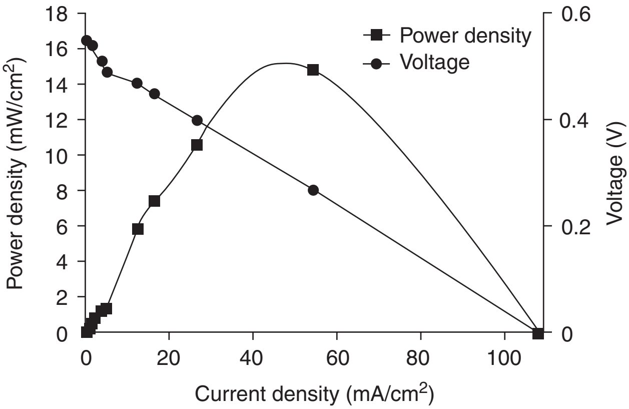
(a) the growing of a thermal oxide used to be used as a sacrificial layer;
(b) the growing of a polysilicon layer by low-pressure chemical vapor (LPCV) deposition to support the MEA;
(c) the patterning and removing of polysilicon by reactive ion etching (RIE);
(d) the opening of input and output holes for hydrogen feeding;
(e) the etching of oxide to form the reaction chamber (Schilp, 2001);
(f) the deposition of a Pt layer to form the current collector.


Membrane orifice
Etched area
Pillar

15.18 Detail of mask patterning for pillars design.


(a)
(b)
(c)
(d)
(e)
(f)
■ Platinum
■ Polysilicon
■ Oxide
□ Silicon

15.19 (a–f) Fabrication process for PEM silicon fuel cell. See text for explanation of steps.



15.20 Measurement of power density and output voltage generated by the PEM silicon fuel cell for a 4-mL/min H_2 flow rate. The figure shows the peak power of 15 mW/cm^2, under about 57-mA/cm^2 current load.

All the steps for this process use standard of MEMS devices manufacturing techniques, which make the industrial-level production easier and more cost-effective.

Experimental results can be seen in Fig. 15.20. A peak power density of 15 mW/cm^2 was obtained for a 4 mL/min H_2 flow rate at 1 atm. The maximum output voltage was 560 mV. Experimental tests were carried out at an ambient temperature of 25°C. The hydrogen flow rate of 4 mL/min was chosen in order to maximize generated power density.

15.5 Applications of micro-fuel cells

The initial applications of fuel cells were very diverse. Research teams were developing different electrolyte technologies and the resulting fuel cells were applied, among other applications, in power generation for NASA space missions. Hydrogen and oxygen were used to produce electric power and water – products that were very much appreciated in space vehicles. The typical power of such fuel cells ranged between 5 and 10 kW. Larger fuel-cell systems were developed in co-generation plant to produce both electric power and heat in domestic applications. In these cases, power was increased to an order of magnitude of 200–400 kW.

The use of fuel cells in portable devices first appeared in military applications. The necessity of personal communication systems with sufficient autonomy lead to the development of micro-fuel cells as power sources. The direct methanol fuel cell (DMFC) is the most suitable for personal applications, because of its low operating temperature and the type of fuel used. DM fuel cells do not use corrosive liquids and have a very simple structure.

Jenny 600S is a small tactical DM fuel cell made by SFC Energy (http://www.sfc-defense.com); it weighs 1.6 kg and delivers up to 25 W in a 10–30 V range. Each fuel cartridge contained 35 cc of methanol, a very safe and suitable fuel in portable applications, with an energy capacity of 400 Wh.

In the area of civil applications, several companies are introducing fuel-cell technology in small devices like cell phones, MP3s and laptop computers. For instance, the Samsung Q35 notebook has been adapted to use DM fuel-cell technology, running for 10–15 h with a 100 cc methanol cartridge. Also, Toshiba has developed a prototype fuel cell for laptops that is capable of generating 20–100 watts of power.

Main manufacturers of electronic power systems are developing fuel-cell-based power sources. Some companies such as Mechanical Technology Inc. (MTI), Motorola, Fujitsu or Samsung SDI are developing rechargeable DM fuel cells for mobile applications. Most of them use methanol for fuel, delivering between 100 mW and 20 W with a capacity about 100 mL (Arunabha, 2007).

15.6 Smart energy management with sun sensors

For decades, considerable effort has been devoted to research into the possibility of obtaining electrical energy in large power plants directly from sunlight by means of thermal concentration techniques. Solar collector and parabolic trough technologies have both had to overcome major technological problems to become efficient and profitable. Among them are the problems related to focusing the maximum radiation on the target; a minor deviation in alignment with the target causes a dramatic drop in system performance.

On the other hand, photovoltaic technology has been well known for a considerable time. The energy efficiency of electro-optic direct conversion is very high, but the use of static structures, in order to reduce installation and maintenance costs, makes collecting the maximum energy throughout the day very difficult.

Both types of system need a solar tracking technique to improve their performance. However, while the use of this technique is mandatory for concentrating solar power plant, for a photovoltaic power plant it is only optional.

Analytical methods have been used to predict the position of the sun. For instance, Mitchaltsky equations provide an analytical curve that describes the path followed by the sun in the sky; an established global positioning system (GPS) coordinates with the date and time on Earth. These methods are quite accurate (error $< 0.1°$); however, they require a relatively complex implementation and periodic recalibration of each of the tracking structures.

To reduce the frequency of this recalibration, robust and precise mechanical structures and engines are used, increasing the price of the system.

The use of sun sensors is an alternative to analytical methods for predicting the position of the sun. These sensors provide a direct measure of position, allowing closure of the control loop of the tracking systems. Many solar power companies include some kind of optical sensor in its parabolic trough collectors in order to track the solar position throughout the day. Fichtner Solar GmbH (www.fichtnersolar.com) uses a very simple photovoltaic two-cell structure in its 150 MW solar power plant in Kuraymat (Egypt). Flagsol GmbH (www.flagsol.com) has designed and engineered the solar field and heat transfer system for the AndaSol 1, 2 and 3 solar thermal power plant in Andalusia (Spain). In the AndaSol plant, the parabolic trough collectors track the sun from east to west using a high-precision optical sensor, thus collecting the maximum solar radiation. Bright Green Energy Ltd (www.wirefreedirect.com) uses a combination of optical solid-states and sun sensors to align the modules to the arc of the sun and measure the solar radiation falling on the array.

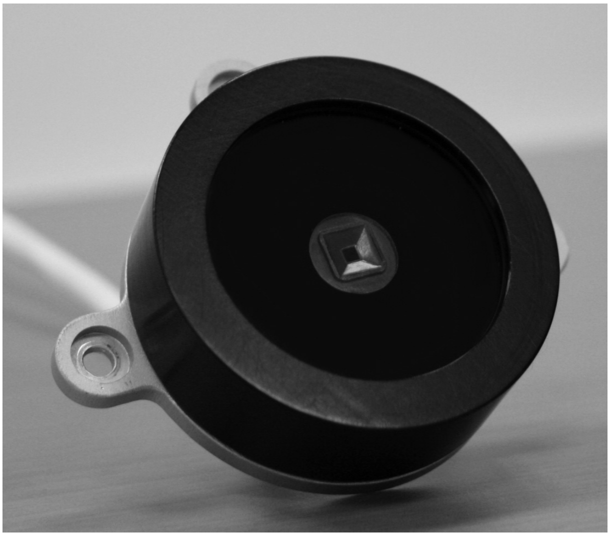
Although the performance achieved by solar sensors is quite good (<0.1°), so far they have been realized with conventional electromechanical technologies, obtaining large devices that are unreliable and expensive. That is why sun sensors have not been widely used in this kind of system yet. However, the use of MEMS technology in this type of sensor solves these problems, so they are now being incorporated into the new tracking systems. Companies such as Solar MEMS Technologies (www.solar-mems.com) have developed a MEMS-based sun sensor for industrial applications (see Fig. 15.21) which is being incorporated into tracker structures like parabolic trough and parabolic dish concentrators.

Moreover, sun sensors are being used in illumination applications to supply natural light to buildings. They are included in two-axes motorized mirrors to focus the solar beam in a fixed place on the building. Sun sensors allow closure of the illumination loop, providing a system with greater accuracy and flexibility and making installation and maintenance more straightforward.

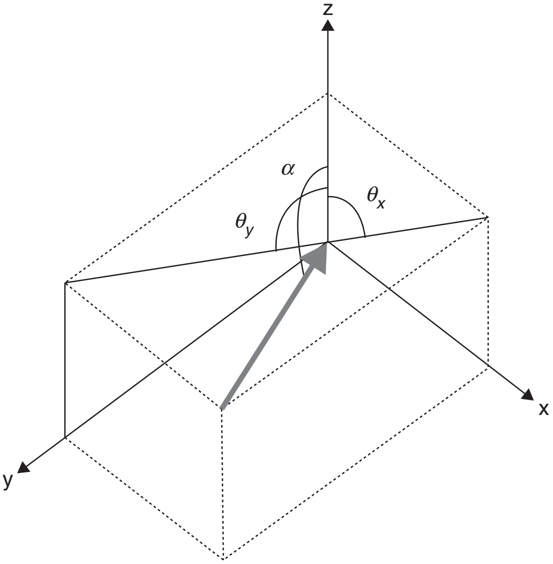
15.6.1 Principles and structure of a sun sensor

The objective of a sun sensor is to measure the angle of incidence of sunshine radiation; in other words, the relative position of the sun in the sky. The relationship between the real angle of incidence α and its components, θ_x and θ_y, in the axes x and y (see Fig. 15.22), is given by Equation [15.13].

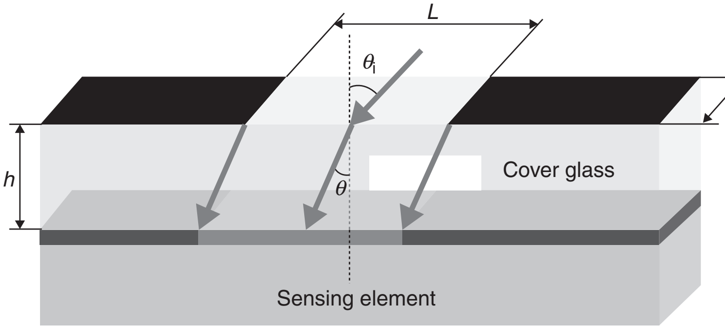
$$\tan^2(\alpha) = \tan^2(\theta_x)\tan^2(\theta_y) \qquad [15.13]$$



15.21 MEMS-based sun sensor, ISS-DX, for industrial applications. (*Source*: Solar MEMS Technologies.)


z
α
θ_x
θ_y
x
y

15.22 Decomposition of a real sunlight beam in both *x* and *y* components.



15.23 Operation principle of a sun sensor.

To measure these angles, it is necessary that the beam of sunlight crosses a window and a cavity to reach a sensing element. From the relative position of the illuminated area in the sensing element, the incidence angle of radiation can be calculated (see Fig. 15.23). The angular range of the sky that can be measured by a sun sensor is known as the ‘field of view’ (FOV).

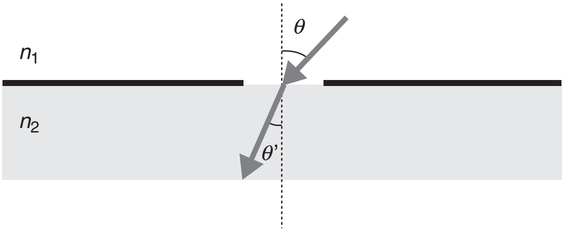
The top window can be manufactured in a variety of ways. One of the most widely used procedures involves drilling a hole in the material that fits over the sensing element. Alternatively, the window can be obtained by metallization and patterning processes on the rear surface of a glass cover which is then bonded to the silicon die.

Realizing the window by drilling a hole in the sensing element cover leaves an empty cavity between the window and the silicon. Thus, independence between axes is guaranteed. In other words, variations on a single axis do not affect the measurement made in the orthogonal axis; the system is therefore much more linear and the data processing simpler. On the other hand, the silicon surface is exposed to external conditions, either atmospheric weather (for terrestrial applications) or dangerous radiation (for space applications).

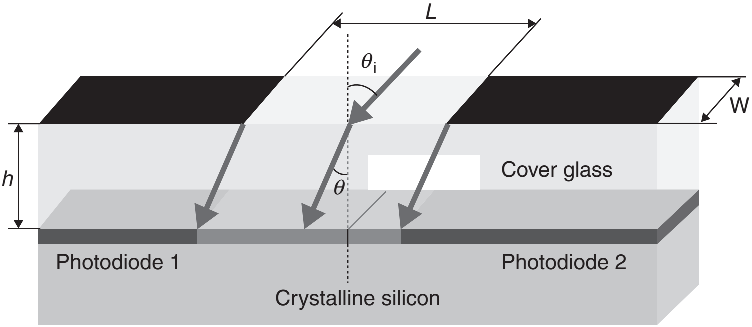
By metallization and patterning processes on the rear surface of a glass cover, the silicon die is protected; the device is therefore more robust, giving it a longer life span and allowing its use in more hazardous environmental conditions, . However, using glass in the window involves a change of material in the path of the light beam, bringing Snell’s law into play.

$$n_1 \sin(\theta) = n_2 \sin(\theta') \qquad [15.14]$$

where n_1 and n_2 are the refractive indices of air (or vacuum) and glass, respectively, θ is the angle of incidence and θ' the refracted angle (see Fig. 15.24).



15.24 Refraction of a light beam crossing an interface between two different materials.



15.25 Basic structure of an analog sun sensor.

The main drawback to this interface is that the independence of the two angles is lost. The decomposition of both original and refracted angles produces different x and y components, due to Snell's law. This means that the variation of the illuminated area in the silicon die, when moving in a single axis, also depends on the angular position on the orthogonal axis.

The way to process this information and obtain accurate measurements depends on the type of technology used. Traditionally, sun sensors are classified as either digital or analog.

Analog sun sensor

The simplest analog structure of a sun sensor consists of a pair of silicon photodiodes and an opaque surface with a ($L \times W$ μm^2) window placed at h μm height (see Fig. 15.25). The light beam enters the window and illuminates both photodiodes, the areas illuminated being dependent on the angle of the beam. Each photodiode will generate a current proportional to the illuminated area, which will be measured by the data acquisition system (Ortega, 2010).

Using this structure, it is possible to measure the angle of incidence on a single axis. To obtain the angle of incidence in two axes (and therefore precisely locate the light source), it is necessary to place two such structures orthogonally.

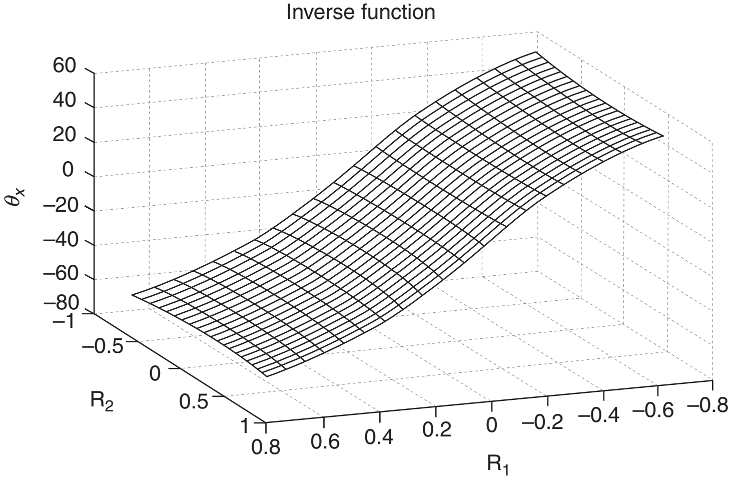
As Equation [15.15] shows, by using a differential measurement of these currents a dimensionless function R_i is obtained. It depends on the angle of incidence and is independent of factors that affect all the photodiodes equally (intensity of light radiation, temperature, etc.).

$$R_i = \frac{v_1 - v_2}{v_1 + v_2} \qquad [15.15]$$

By combining the functions of both axes, the characteristic curves of the sensor $\theta_x = f(R_1, R_2)$ and $\theta_y = f(R_1, R_2)$ are obtained. During normal operation, the procedure consists in measuring the voltage and calculating the corresponding R functions, and then applying them to the characteristic curve in order to obtain the value of the angular position (see Fig. 15.26).

Digital sun sensor

In a digital sensor, the sensing element is either a charge-coupled device (CCD) or active pixel sensor (APS) image detector onto which the beam of sunlight is projected. The small dimension of the pixels means that the projection of the beam of sunlight illuminates a set of them (Liebe, 2001; Mobasser, 2003). Using image processing algorithms, it is possible to obtain



15.26 Angle θ_x as a function of the sensor's responses R1 and R2.

Table 15.5 Usual accuracy and field of view (FOV) of digital and analog sun sensors

	FOV	Accuracy	
		Analog	Digital
CSS	±60°	1°	0.1°
FSS	±5°	0.01°	0.005°

the centroid of this projection – and, hence, the angle of incidence – with great accuracy (Enright, 2008).

The most common digital design consists of a silicon die coated with a thin layer of chrome and a layer of gold with hundreds of small pinholes. This structure is placed on top of the image detector at a distance of less than 1 mm. Images of the sun are formed on the detector when the sun illuminates the assembly. Software algorithms must be able to identify the individual pinholes on the image detector and calculate the angle to the sun (Mobasser, 2001). The accuracy of this kind of sensor depends on several factors, including the pattern implemented in the mask, the holes (number, size and separation between them) and the algorithms used to obtain the angle of incidence.

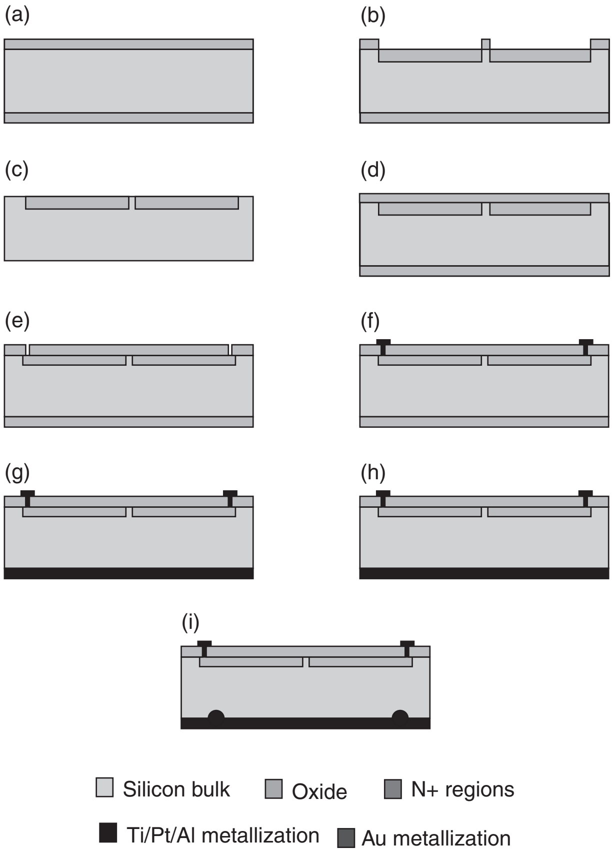
Sun sensors are classified into two groups according to the accuracy of their sensor: coarse sun sensors, which have a large FOV but a low degree of accuracy; and fine sun sensors, which provide a high degree of accuracy from a small FOV. In a complete detector system, the best performance is obtained by using a combination of both kinds of sensor. A comparison of typical values of accuracy for coarse and fine sun sensors is shown in Table 15.5.

To summarize, an analog sensor provides robustness and simple signal processing, whereas a digital sensor is more sensitive to failure and has a more complex signal processing. However, the accuracy achieved with the digital sensors is greater than provided by analog sensors.

15.6.2 Analog sun sensor manufacturing

The analog design is simple and uses large photodiodes (about 1 or 2 mm^2), so analog sun sensors are very robust against environmental conditions, making them reliable as a navigation instrument in space applications.

Among the topologies developed by different research groups, there are sensors that consist of two basic structures arranged orthogonally (Ortega, 2010), an advanced version with the FOV divided into sectors (Delgado, 2010), a four-quadrant sensor in which the axes are identified by processing the voltages in pairs (Maqsood, 2010) or a version with triangular-shaped



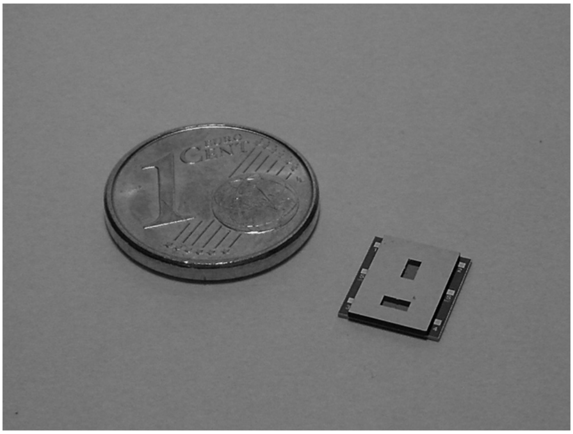
15.27 (a–i) Fabrication process of a two-photodiodes sun sensor. See text for explanation of steps.

photodiodes (Pedersen, 2003). An analog design with a different concept was developed by Bohnke (2005), where the sensor is composed of a semi-spherical photodetector element. The current is generated and distributed between different electrodes. The distribution of the photocurrent to the electrodes depends on the angle of incidence of the sunlight.

The fabrication of each of these types of sensor is similar and can be summarized in three main steps: silicon die fabrication, cover glass metallization and the bonding process (the latter two, only if the sensor has a layer of glass). The silicon die is fabricated using p-type crystalline silicon wafers.

The fabrication process of the sun sensor developed by Ortega (2010) consists of the following steps, and can be seen in Fig. 15.27:

(a) thermal oxidation;
(b) definition of the emitter regions (N+ regions) of the photodiodes using standard photolithography;



15.28 Prototype of an analog sun sensor.

(c) SiO_2 etching at the front and back surfaces;
(d) SiO_2 passivation;
(e) definition of the emitter contacts in N+ regions by photolithography and SiO_2 etching;
(f) metallization using sputtering and the lift-off technique;
(g) Al metallization on the rear surface and thermal annealing;
(h) metallization in the pads and thermal annealing;
(i) laser-fired contacts in the back surface.

The cover glass metallization is made as follows:

- photoresist deposition on the front and lithography;
- Al metallization using the sputtering technique;
- metal patterning using the lift-off technique.

Finally, a transparent and non-conductive epoxy resin is deposited carefully on the silicon die in four peripheral points (to avoid covering the electrodes and windows) in order to bond the silicon and glass cover. Figure 15.28 shows a prototype of the analog sun sensor developed by Ortega (2010).

15.6.3 Sun sensor strategy for energy management

A sun sensor is used in either of the following modes of operation:

- full operating range (full FOV);
- around a reference point, usually the origin (0.0).

Like most sensors, sun sensors can be used merely as an instrument for measuring the sun's position in a given range or they can be incorporated into a tracking system as an active element. When incorporated into a tracking system, the sensor usually works around the origin (0.0).

Sensor requirements are different for each of these applications. In the full operating range, one or more sensors with wide FOV (eg + –60°) is necessary to cover the full angular field where the sun position has to be known. The system operates in an open loop and the required accuracy is high throughout the operating range. To achieve this precision, the sensor has to be calibrated before it is deployed. In this way, manufacturing tolerances are compensated and no unacceptable errors occur in the measurement function.

Good calibration requires the use of a solar simulator and a positioning system, which range throughout the FOV and measure the response of this particular sensor in that range. This information should be included in a memory device so that it can be used later by a microprocessor that is responsible for performing the positional calculations. Both the microprocessor and the memory can be included in the device itself, as in the case of a digital sensor.

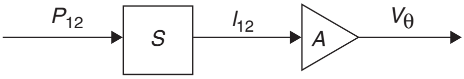
The relationship between the difference in optical power (P_{12}) received by photodiodes 1 and 2 of the sensitive element and the difference in currents generated by both of them (I_{12}) is determined by the function $I_{12} = f(P_{12})$. If the sensor operates close to the origin of reference, both differences are close to zero ($P_{12} \sim 0$), and the function f can be approximated, as shown by Equation [15.16]:

$$I_{12} = 0 + \frac{\partial f}{\partial P_{12}} P_{12} = SP_{12} \qquad [15.16]$$

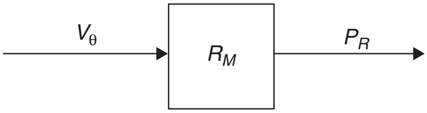
Equation [15.15] shows that the differences of incident light and electrical current (P_{12} and I_{12}) are related by a coefficient known as sensitivity S.

Figure 15.29 shows the equivalent model of the sensitive element (S) whereby the incident optical power (P_{12}) is converted into electricity (I_{12}). Using an electronic process, represented by A, the angle of incidence of radiation is given as a voltage V. This electronic process includes both the current–voltage conversion and signal amplification stages.

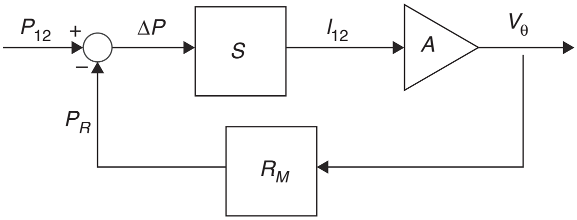
A tracking system includes a set of motors that are driven by a control signal V moving the support structure, which causes a variation in the difference



15.29 Equivalent model of a sensitive element.



15.30 Equivalent model of an actuator system.



15.31 Equivalent model of a closed loop tracking system.

in optical power received by the photodiodes (PR). This operating mode is outlined in the Fig. 15.30.

Including a solar tracking system, we can achieve a sensor that works around the origin (0.0). The system operation is shown in the Fig. 15.31.

The relationships of different variables of the system are described by:

$$V_\theta = AI_{12} \qquad [15.17]$$

$$I_{12} = S\Delta P \qquad [15.18]$$

$$\Delta P = P_{12} - P_R \qquad [15.19]$$

$$P_R = R_M - V_\theta \qquad [15.20]$$

where of

$$V_\theta = \mathrm{AS}\left(P_{12} - R_M V_\theta\right) \qquad [15.21]$$

$$\left(1 - \mathrm{ASR}_M\right)V_\theta = \mathrm{ASP}_{12} \qquad [15.22]$$

$$V_\theta = \frac{\mathrm{ASP}_{12}}{\left(1 - \mathrm{ASR}_M\right)} \qquad [15.23]$$

Supposing that $\mathrm{ASR}_M >> 1$, then

$$V_\theta = \frac{P_{12}}{R_M} \tag{15.24}$$

In conclusion, when the sun sensor works in a closed loop:

- the scale factor is determined by the transducer, which is usually very linear, precise and repetitive;
- the nonlinearity of the sensor (*f*) is removed, because the sensor is working at all times around 0.0, so higher accuracy will be achieved.

The requirements of sun sensors used in this operating mode are typically those of a fine sun sensor, where the FOV and accuracy are usually less than ± 15° and 0.1°, respectively. The use of fine sun sensors around the origin (0.0), where accuracy is even greater, maximizes the performance of photovoltaic energy conversion by using tracking systems. In addition, a narrow FOV improves the problem of albedo. Although a fine sun sensor is the best option for the closed loop control of a tracking system in transient situations, such as at sunrise or in partly cloudy weather, the sensor may be out of the FOV. In this situation, a coarse sun sensor works more robustly due to its wider FOV. A combination of both types is usually employed.

15.7 Conclusion

In this chapter, we showed photovoltaic and fuel-cell technologies as a suitable option for powering MEMS for smart energy management, such as bio-implantable devices or portable electronic devices.

We introduced the basic working principles of photovoltaics and followed this by describing the fabrication processes of different photovoltaic minimodule approaches based on silicon: multichip module (MCM), silicon-on-insulator (SOI) and fusion-bonding technologies.

New photovoltaic devices and sensors are emerging continuously. The materials, structures and strategies used have a fundamental role in achieving a successful application.

15.8 References

Al-sarawi, S.F., Abbot, D. and Franzon, P.D. (1998) 'A review of 3-D packaging technology'. *IEEE Transactions on Components, Packaging, and Manufacturing Technology-Part B*, **21**, 2–14.

Andò, B., Baglio, S., Trigona, C., Dumas, N., Latorre, L. and Nouet, P. (2010) 'Nonlinear mechanism in MEMS devices for energy harvesting applications'. *Journal of Micromechanics and Microengineering,* **20**, 125020–125032.

Arunabha, K., Jang, J.H. Gil, J.H. Jung, C.R. Lee, H.R. Kim, S.-H. Ku, B. and Oh, Y.S. (2007) 'Micro-fuel cells – Current development and applications'. *Journal of Power Sources*, **170**(1), 67–78.

Beaumont, B., Guillaume, J.C., Vileia, M.F., Salètes, A. and Vèrié, C. (1991) 'High efficiency conversion of laser energy and its application to optical power transmission', *Proceedings of 22nd IEEE Photovoltaic Specialist Conference*, Las Vegas, USA, October 1991, 1503–1507.

Bermejo, S., Ortega, P. and Castañer, L. (2005) 'Fabrication of monolithic photovoltaic arrays on crystalline silicon by wafer bonding and deep etching techniques'. *Progress in Photovoltaics: Research and Applications*, **13**, 1–9.

Bermejo, S. and Castañer, L. (2005) 'Dynamics of MEMS electrostatic driving using a photovoltaic source'. *Sensors and Actuators A: Physical*, **121**(1), 237–242.

Bohnke, T. and Stenmark, L. (2005) 'Development of a MOEMS sun sensor for space applications'. *Digest of Technical Papers* (*TRANSDUCERS '05*). 13th International Conference on Solid-State Sensors, Actuators and Microsystems, vol.1, 589–592.

Brendel, R. and Auer, R. (2001) 'Photovoltaic mini-modules from layer transfer by the porous silicon process'. *Progress in Photovoltaics: Research and Applications Progress in Photovoltaics: Research and Applications*, **9**, 439–443.

Castañer, L., Pons, J., Nadal-Guardia, R. and Rodríguez, A. (2001) 'Analysis of the extended operation range of electrostatic actuators by current-pulse drive'. *Sensors and Actuators: Physical A*, **90**(3), 181–190.

Castañer, L.and Senturia, S.D. (1999) 'Speed-energy optimization of electrostatic actuators based on pull-in'. *IEEE Journal of Microelectromechanical Systems (JMEMS)*, **8**(3), 290–298.

Castañer, L., Rodriguez, A., Pons, J. and Senturia, S. D. (1999) 'Measurement of power-speed product of electrostatic actuators', *Transducers '99*, Sendai, Japan, 7–10 June, 1772–1775.

Cook-Chennault, K.A., Thambi, N. and Sastry, A. (2008) ' Powering MEMS portable devices – a review of non-regenerative and regenerative power supply systems with special emphasis on piezoelectric energy harvesting systems'. *Smart Materials and Structures*, **17**, 1–33: 10.1088/0964-1726/17/4/043001.

Dehlinger, P.J. (2000) 'Powering biomedical microdevices'. *Biomedical Microdevices, An International Journal*, **2**, 9–10.

Delgado, F.J., Ortega, P., Tarrida, C.L., García, J., Angulo, M. and Quero, J.M. (2010) 'A new design of high precision solar microsensor for satellite applications', *Proceedings of the 9th IEEE Conference on Sensors*, Waikoloa, Hawaii, 1–4 November 2010.

Dinesh, B., Sweeti, B., Sanat, G. and Manoj, J. (2010) 'Pacemakers charging using body energy. *Journal of Pharmacy and BioAllied Science'*, **2**(1), 51–54.

Enright, J.P. and Godard, (2008) 'Design optimization of a digital sun sensor for use with parametric processing'. *IEEE Transactions on Instrumentation and Measurement*, **57**(10), 2188–2195.

Keller, S., Scheibenstock, S., Fath, P., Willeke, G. and Bucher, E. (2000) 'Theoretical and experimental behavior of monolithically integrated crystalline silicon solar cells'. *Journal of Applied Physics*, **87**, 1556–1562.

Kelley, S.C., Deluga, G.A. and Smyrl, W.H. (2002) 'Miniature fuel cells fabricated on silicon substrates'. *AIChE Journal*, **48**(5), 1071–1082.

Koeneman, P.B., Busch-Vishniac, I.J. and Wood, K.L. (1997) 'Feasibility of micro power supplies for MEMS'. *Journal of Microelectromechanical Systems*, **6**, 355–362.

Lee, J.B, Chen, Z., Allen, M.G., Rohatji, A. and Arya R. (1995) 'A miniaturized high voltage solar cell array as an electrostatic MEMS power supply'. *Journal of Microelectromechanical Systems*, **4**, 102–108.

Liebe, C.C. and Mobasser, S. (2001) 'MEMS based sun sensor', *IEEE Proceedings on Aerospace Conference*, **3**, 1565–1572.

Liu, X., Zhang, B., Zhang, Y., He, H., Li, J., Wang, S., Yuan, Z. and Deng, H. (2010) 'Development and characterization of a novel air-breathing micro direct methanol fuel cell stack for portable applications'. *Journal of Micromechanics and Microengineering,* **20**: 10.1088/0960-1317/20/10/104008.

Luque, A., Moreno, J.M., Brey, J., Ellis, C.D., Quero, J.M. and Wilamowski, B.M. (2010) 'Monolithically integrated gas distribution chamber for silicon MEMS fuel cells'. *Journal of Microelectromechanical Systems*, **19**(2), 384–390.

Maqsood, I. and Akram, T. (2010) 'Development of a low cost sun sensor using quad-photodiode', Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION, Indian Wells/Palm Springs, California USA.

Mobasser, S., Liebe, C.C. and Howard, A. (2001) 'Fuzzy image processing in sun sensor'. *10th IEEE International Conference on Fuzzy Systems,* Melbourne, Australia, vol. 3, 1337–1342.

Mobasser, S. and Liebe, C.C. (2003) 'MEMS based sun sensor on a chip'. *Proceedings of 2003 IEEE Conference on Control Applications 2003* (*CCA 2003*), vol. 2, 1483–1487.

Muanghlua, R., Cheirsirikul, S. and Supadech, S. (2000) 'The study of silicon thermopile'. *Proceedings of the International TENCON*, Kuala Lumpur, Malaysia, 24–27 September 2000, vol. 3, 226–229.

Nathanson, H.C., Nevel, W.E., Wickstrom, R.A. and Davis Jr. J.R. (1967) 'The resonant gate transistor'. *IEEE Transactions Electron Devices*, ED-**14(**3), 117–133.

Nisar, A., Afzulpurkar, N., Mahaisavariya, B. and Tuantranont, A. (2008) 'MEMS-based micropumps in drug delivery and biomedical applications'. *Sensors and Actuators B,* **130**, 917–942.

Ortega, P., Bermejo, S. and Castañer, L. (2001) 'Multichip module photovoltaic mini-arrays'. *IEEE Transactions on Advanced Packaging*, **24**, 169–174.

Ortega, P., Bermejo, S., Cabruja, E. and Castañer, L. (2002) 'Small PV generators assembled using multichip module technology' Proceedings of the 6th European Space Power Conference, Oporto, Portugal, 6–11 May 2002, 545–550.

Ortega, P., Bermejo, S. and Castañer, L. (2008) 'High voltage photovoltaic mini-modules'. *Progress in Photovoltaics: Research and Applications*, **16**, 369–377.

Ortega, P., Lopez, G., Ricart, J., Dominguez, M., Castañer, L., Quero, J.M., Tarrida, C.L. and García, J. (2010) 'Fabrication of a miniaturized two axis sun sensor for satellite applications'. *IEEE Sensors Journal*, **10**(10), 1623–1632.

Pedersen, M., Hales, J. H. and Fléron, R. W. (2003) 'Linear two-axis MOEMS sun sensor and the need for MEMS in space'. *Proceedings of the 54th International Astronautical Congress*, Bremen, Germany, 29 September–October 2003.

Radbill, A.E., Triedman, J.K., Berul, C.I., Fynn-Thompson, F., Atallah J., Alexander, M.E., Walsh, E.P. and Cecchin, F. (2010) 'System survival of nontransvenous implantable cardioverter-defibrillators compared to transvenous implantable cardioverter-defibrillators in pediatric and congenital heart disease patients'. *Heart Rhythm*, **7**(2), 193–198.

Sari, I., Balkan, T. and Külah, H. (2010) 'An electromagnetic micro power generator for low-frequency environmental vibrations based on the frequency upconversion technique'. *Journal of Microelectromechanical Systems*, **19**, 14–27.

Schilp, A., Hausner, M., Puech, M., Launay, N., Karagoezoglu, H. and Laermer, F. (2001) 'Advanced etch tool for high etch rate deep reactive ion etching in silicon micromachining production environment', *Proceedings of MST*, 425–430.

Seeger, J.I. and Crary, S.B. (1997) 'Stabilization of electrostatically actuated mechanical devices'. *Transducers '97*, Chicago, USA, 16–19 June, 1133–1136.

Shreve, K.P., DiNetta, L.C., Cotter, J.E., Bower, J.R., Ruffins, T.R. and Barnett, A.M. (1996) 'Initial results for the silicon monolithically interconnected solar cell product' IEEE 25th Photovoltaic Specialist Conference, Washington, DC, 13–17 May, 665–668.

Sodano, H.A., Inman, D.J. and Park, G. (2004) 'A review of power harvesting from vibration using piezo electric materials'. *Shock and Vibration Digest*, **36**, 197–205.

Van Kerschaver, E. and Beaucarne G. (2006) 'Back-contact solar cells: a review'. *Progress in Photovoltaics: Research and Applications*, **14**, 107–123.

16

Radio frequency (RF)-MEMS for smart communication microsystems

D. DUBUC and K. GRENIER, LAAS-CNRS,
University of Toulouse, France

DOI: 10.1533/9780857099297.2.472

Abstract: This chapter sets out the applications of radio frequency micro-electro-mechanical systems (RF-MEMS) technologies for smart communication microsystems. It begins with a brief presentation of RF-MEMS technology and associated components, and then offers a detailed overview of RF-MEMS-based reconfigurable circuits, paying specifical attention to the reliability and power capability of RF-MEMS components. The chapter closes with a discussion of the co-integration of RF-MEMS circuits with integrated circuits (ICs) leading to smart communications microsystems is presented and global conclusions are offered.

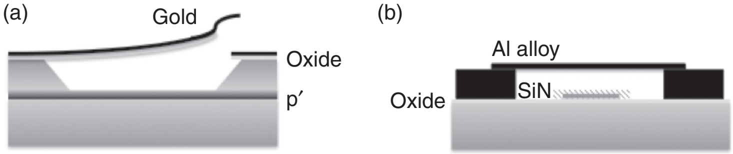
Key words: radio frequency micro-electro-mechanical systems (RF-MEMS), microsystems, microwave.

16.1 Introduction

Joseph Henry invented the first electromechanical device, called a 'switch' or 'relay', in 1835. It was composed of a bulky electromagnet, which activated an armature that contacted two electrodes. That relay was the brick of the first computer in the late nineteenth century. The clock speed was unfortunately limited to tens of Hertz, which motivated the well-known extensive use of vacuum tubes, followed by transistors during the twentieth century. After several decades of semiconductor foundry developments, complementary metal oxide semiconductor (CMOS) chips are now drastically reduced in size, operate at GHz frequencies and consume very low power.

Is the relay definitely buried?

In 1979, K.E. Petersen developed the first 'micromechanical membrane switches' that 'fill the gap between conventional silicon transistors and mechanical electromagnetic relays' (Petersen, 1979). This was the first starting point, which opened the route to micro-electro-mechanical systems (MEMS). The key idea, an original concept at that time, was to merge silicon-



16.1 RF MEMS switches. (a) Petersen (1979) and (b) Goldsmith *et al.* (1995).

based microfabrication capabilities with the mechanical relay approach. The proposed MEMS-based electrical switch exploited electrostatic force applied between a p-doped silicon layer and a bi-material (SiO_2 and gold) membrane in order to pull-down a movable structure and perform an electrical contact (Fig. 16.1a). The miniaturization achieved by the microfabrication of such a relay lead to the envisioning of a matrix-addressable cross-point switching array for AC signals. However, the associated maximum operation frequency (200 kHz) was too low to cover radio frequency (RF) or microwave applications.

During the 1990s, the electrical and microwave community assisted with the emergence of a new class of device: micro-relays capable of processing analog, RF and microwave signals.

In 1990, Halg *et al*. introduced the first integrated micro-electro-mechanical nonvolatile memory cell (Halg, 1990); 1 year later, Larson *et al*. (1991) demonstrated a microwave rotating transmission line switch. Measurements performed up to 45 GHz were already impressive with insertion losses lower than 0.5 dB and isolation greater than 35 dB. Several years later, in 1995, Goldsmith *et al*. introduced what was to become one of the best-known bridge configurations for RF-MEMS switches, with a capacitive contact, as illustrated in Fig. 16.1b. The era of RF-MEMS was emerging apace (Goldsmith *et al*., 1996; Pacheco *et al*., 1998; Muldavin and Rebeiz, 1999; Sovero 1999; Yao *et al*., 1999; Rebeiz *et al*., 2003). After more than 15 years, one can wonder: what potential do MEMS hold for RF and microwave applications? What are the bottlenecks in this technology? Before answering these questions, we present the basics of RF-MEMS technology.

16.2 RF-MEMS technology and devices

The concept and technology of RF-MEMS rely on bulk and surface micromachining techniques, which have already proven their potentialities in high-volume commercial applications such as the micro-mirrors in video projectors, printer inkjet heads and, more generally, for integrated sensors (the accelerometer, for instance) and actuators.

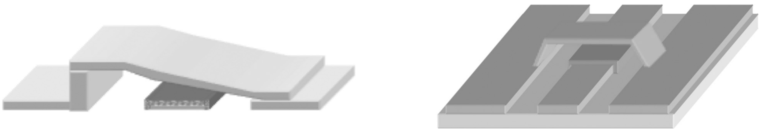
Tunable RF-MEMS devices typically employ bridge or cantilever configurations, as shown in Fig. 16.2. The movable membrane exhibits a vertical movement during actuation. Some RF-MEMS may exploit lateral movement; for example, the switch developed by Oberhammer *et al.* (2006), which involves the superposition of two bonded substrates (silicon and glass wafers). The top silicon substrate is dry etched to realize the moveable component, whereas the bottom glass wafer is wet etched to allow the free movement of the metalized silicon beam.

Different types of actuation have also been studied and most of the RF-MEMS architectures rely on electrostatic forces because of ease of fabrication and low power consumption. For some specific applications, electrothermal, piezoelectric and magnetostatic actuations may also be encountered – each of which presenting specific advantages and drawbacks.

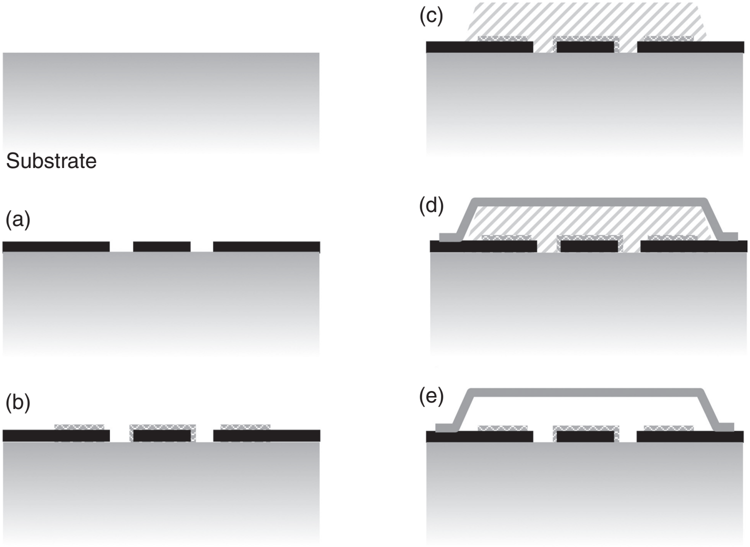
Numerous RF-MEMS device architectures may be found in the literature and Rebeiz (2003) gives an overview. In this chapter, we focus on the devices introduced by Goldsmith *et al.* (1995), in which a mechanical bridge is moved up and down over a coplanar waveguide due to an electrostatic force. When the metallic bridge is down, it contacts a thin dielectric layer; this prevents a DC-contact with the line. We call this contact 'capacitive' in contrast with the ohmicone, when one metal makes contact with another. For this configuration, the process flow is globally divided in five main steps, as illustrated in Fig. 16.3 (Grenier *et al.*, 2004, 2005).

The first step consists of defining the RF transmission line, which is sketched with a coplanar waveguide. In the case of capacitive switches, a dielectric layer is then deposited and patterned on the lines (see Fig. 16.3b). The next step is to elaborate a sacrificial layer (see Fig. 16.3c), which will support the movable membrane during its fabrication (see Fig. 16.3d). Finally, the sacrificial layer is removed to release the RF-MEMS structure.

With this technology, we are able to integrate, with both microwave passive and active circuits, switches and varactors featuring high-level performances. Before giving some examples of architecture in which RF-MEMS



16.2 Configurations of RF MEMS. Bridge configuration (right); cantilever configuration (left).

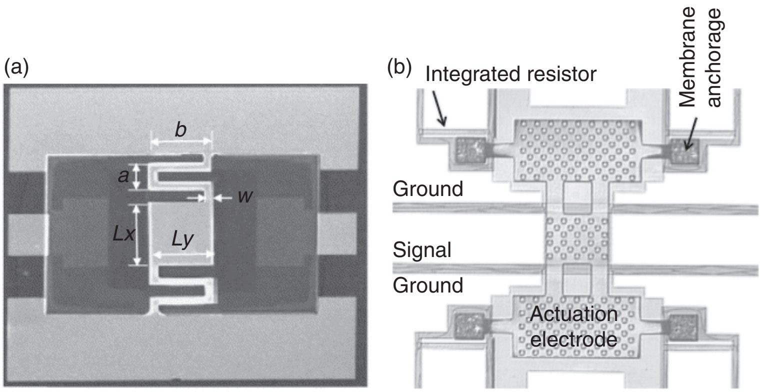


16.3 (a–e) Technological process flow of RF MEMS devices.

devices bring real added value, let us examine achievable performances with such devices.

Figure 16.4a presents high-isolation X-band MEMS capacitive switches (Tang *et al.*, 2005). The pull-down voltage has been optimized around 20 V with the use of a low-effective spring constant of serpentine folded suspensions. Even if this value is rather high compared with the 3.3–1.8 V used for CMOS-devices, it brings about a good compromise between a low operating voltage and a high degree of reliability. Charge pump circuits may also be associated with such RF-MEMS devices in order to operate under low voltage circumstances. This is a minor drawback with respect to the high level of achievable microwave performance. Insertion losses of 0.1–0.35 dB for isolation of 16.5–28 dB are measured for the device presented in Fig. 16.4a, which show a clear improvement compared with solid-state switches (Rebeiz, 2003). Figure 16.5 illustrates the operation of such capacitive RF-MEMS switches. The mobile metallic membrane defined, with the coplanar line underneath, a variable capacitor featuring a low value when the membrane is up, and a high value when the membrane is in contact with the dielectric. A low value of the off-state capacitor assures a low-loss operation as a high on-state capacitor value permits high-isolation performances to be achieved.

Similar devices (Grenier *et al.*, 2005) are presented in Fig. 16.4b and have been characterized from 1 to 40 GHz. The main results consist in an insertion



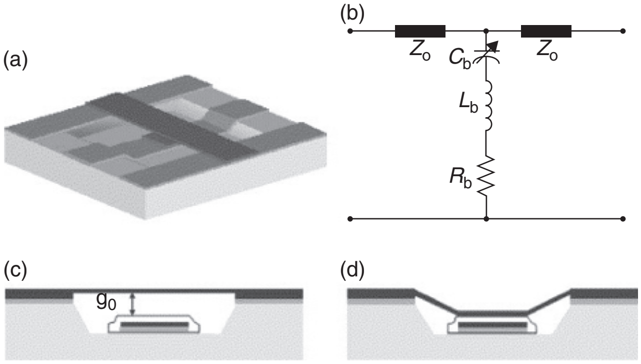
16.4 RF-MEMS capacitive switches. (a) With serpentine folded suspensions and (b) with separate actuation electrodes.

loss lower than 0.2 dB up to 20 GHz (including the CPW ports) and isolation better than –15 dB from 10 to 40 GHz. The up-state (Cup) and down-state (Cdown) capacitances with values of 65 fF and 2 pF, respectively, were extracted from the measurements. It must be pointed out that the Cdown/Cup ratio exhibits a value of 30, which is sufficient to assure its good operation as a switch.

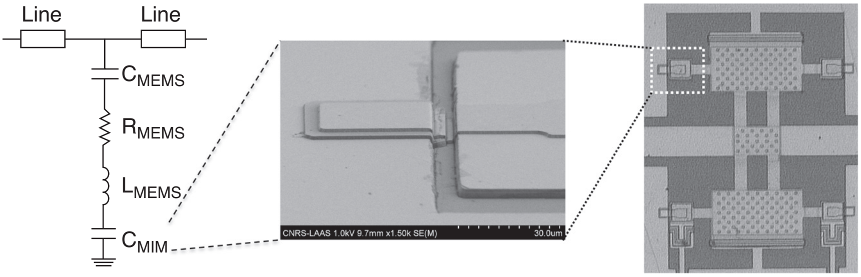
The capacitive RF-MEMS device shown in Fig. 16.4b has been developed through a low-loss technology compatible with an above-IC integration (Grenier *et al*., 2005). It consists of a metallic gold membrane 3 μm high and a 0.25 μm thick silicon nitride- (SiN-)based dielectric. Observe that the attracting electrodes are decoupled from the RF signal in order to facilitate mechanical, electromechanical and microwave design (Ducarouge *et al*., 2004). It should be pointed out that this translates to a longer membrane that necessitates a specific anchor design to ensure the high quality of the contact.

As far as the systems tunability is concerned, varactors featuring a capacitive ratio in the range 2–5:1 are also required (together with RF-MEMS switches) (Shen and Baker, 2005). Consequently, new MEMS configurations which integrate a constant metal-insulator-metal (MIM) capacitor have been developed, as presented in Fig. 16.6. The idea is as simple as it is efficient. Adding a fixed capacitor in series with the MEMS capacitor tends to reduce the resulting range of the total capacitor, as well as the consequent ratio.

The technological process flow depicted in Fig. 16.3 has been modified to integrate these MIM capacitors. Two additional steps are required. After the elaboration of the RF lines, the MIM dielectric (in SiN) is deposited by plasma enhanced chemical vapor deposition (PECVD) and then



16.5 RF-MEMS capacitive switches. (a) Typical schematic, (b) equivalent electrical model, (c) cross-sectional view in the up-state and (d) down-state. (*Source*: Tang *et al*. (2005).)



16.6 RF-MEMS capacitive varactor.

patterned. A top metallization is realized by evaporation and patterned to define the top electrode of the fixed capacitor. The MEMS process restarts then with the deposition of the MEM dielectric and continues until the final release of the structure. Because of certain technological limitations, MIM capacitors have to present a value in the range of between 100 and 200 fF. The MIM capacitor's value corresponds to few 100s fF, which leads to varactor values (MEM and MIM capacitors in serial configuration) of ~100 and ~500 fF in the up and down states, respectively. This results in a capacitive ratio from 2 to 5 (Bordas *et al*., 2008).

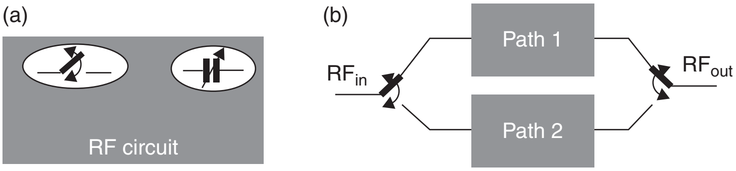
Vähä-Heikkilä *et al*. have proposed another solution based on metal-air-metal (MAM) capacitors with RF-MEMS. It leads to a higher quality factor, as no dielectric losses appear in the MAM device. A 150% improvement in the off-state quality factor, which presents a value of 154, was obtained at 20 GHz (Vähä-Heikkilä and Rebeiz, 2004) with MAM capacitors 100 times larger than MIM capacitors.

16.3 RF-MEMS-based circuits for smart communication microsystems

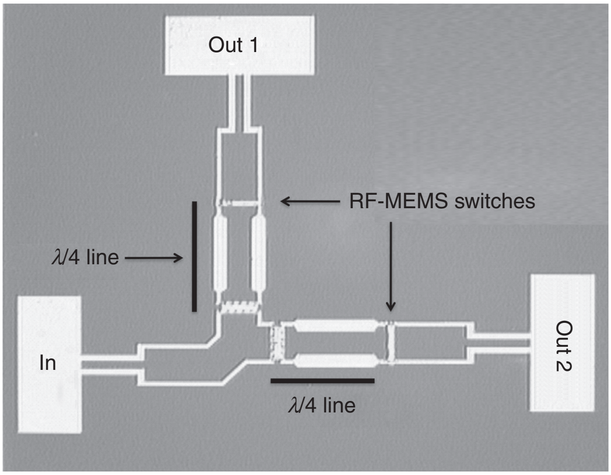
Based on an arsenal of RF-MEMS devices, high performance reconfigurable microwave systems can be envisioned. RF-MEMS devices may, indeed, be added directly to passive functions (e.g. filters, matching network) with no compromise of performance. Basically, there are two ways to bring smartness to circuits: (1) circuit tenability through switches and varactors (Fig. 16.7a) or (2) commutation between two circuits (two RF paths, as illustrated in Fig. 16.7b).

In the case of Fig. 16.7b, the commutation between paths 1 and 2 is performed with two RF-MEMS-based single pole double trough (SPDT) switches (Pacheco *et al.*, 2001). Paths 1 and 2 may correspond to two filters featuring different central frequencies, or to two power amplifiers with distinct nominal operating power, for instance. Figure 16.8 presents an RF-MEMS-based SPDT switch. The SPDT structure comprises two MEMS parallel switches (which permits selection of the active channel), and two quarter wave impedance inverters (which transform the RF short wave emitted by the MEMS switches into an open circuit in order, virtually, to cancel the presence of the inactive channel) (Dubuc *et al.*, 2003). High performances are achieved: only 0.6 dB of insertion losses at 30 GHz are measured, whereas isolation is expected to reach 30 dB. Higher isolation can be reached by cascading RF-MEMS switched as illustrated in Fig. 16.9 (Tang *et al.*, 2005). A 7 dB improvement in the X-band has been measured compared with a single-MEMS switch. This improvement translates into an increase of only 0.3 dB in insertion losses.

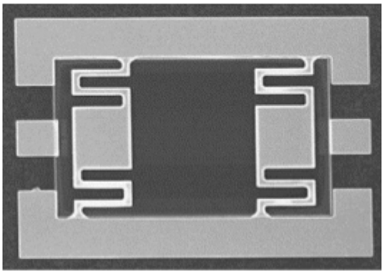
The circuit architectures depicted in Fig. 16.7a cover a very wide range of topologies: two of them, which are probably the most generic architectures, relate to the impedance tuner and phase shifter. These can be used for the intelligent control of the optimum operation of a power amplifier over various frequencies (Qiao *et al.,* 2005). Figure 16.10a depicts a 6-bit impedance tuner operating at 20 GHz. The measurements demonstrate that the tuner is able to match impedance as low as 20 Ω and up to 100 Ω. The corresponding impedance matching ratio of 5:1 is interesting for a wide range of



16.7 RF-MEMS-based smart microsystems. (a) With tunable devices and (b) with switchable circuits.

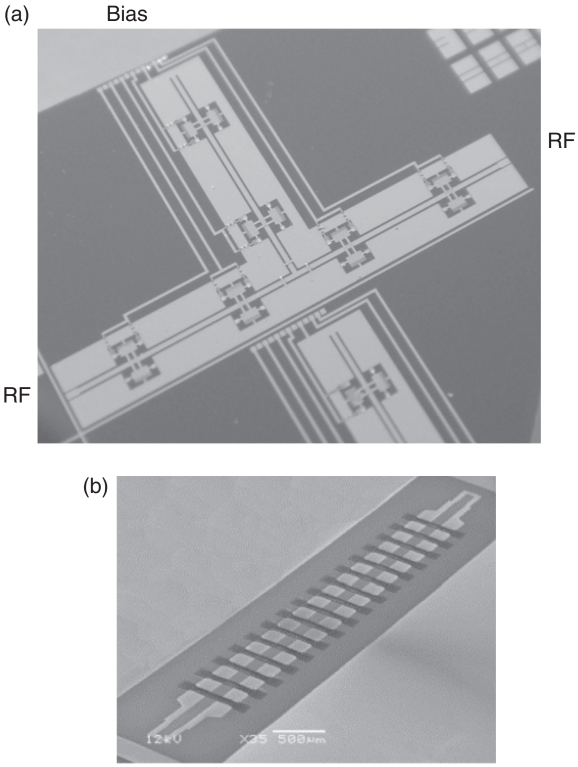


16.8 RF-MEMS-based SPDT.



16.9 RF-MEMS-based high-isolation switch. (*Source*: Tang *et al.* (2005).)

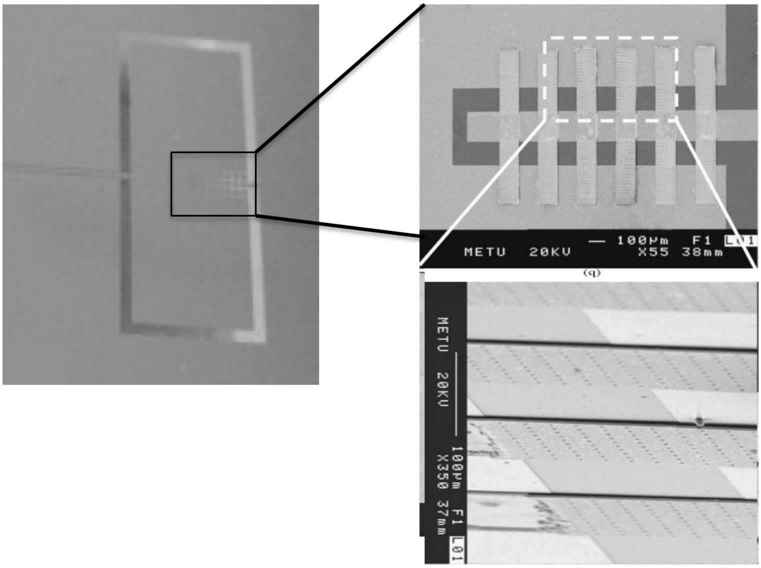
applications, where low noise or power amplifiers and antennas have to be matched under different frequency ranges. The impedance tuner is, then, a key building block of reconfigurable microwave circuits, where impedance matching in various switchable conditions is mandatory. A true time delay (TTD) phase shifter is presented in Fig. 16.10b. It corresponds to MEMS-based switchable capacitors periodically distributed over a coplanar waveguide (Palei *et al.*, 2005). A low level of insertion losses (2.3 dB) for a large phase shift (250°) with a low return loss (–15 dB) has been demonstrated up to the Ku band. This high level of characteristics, induced by the RF-MEMS technology, permits the envisioning of clear performance enhancement in radar and telecommunication systems. Beam steering is probably one of the main applications, for which such circuits may induce breakthrough performances.



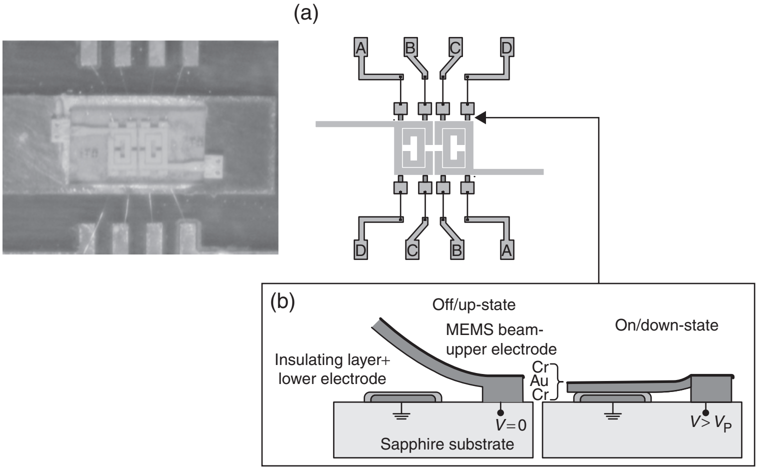
16.10 RF-MEMS-based devices. (a) Impedance tuner and (b) true time delay phase shifter.

The integration of RF-MEMS switches over radiating elements has also been considered in order to achieve smart antennas. Figure 16.11 presents a rectangular patch antenna with a tunable operating frequency, for which MEMS technology is expected to assure high antenna efficiency. A 1 GHz tunability of the antenna resonant frequency has been achieved in the X-band (Topalli *et al*., 2009), with no alteration of the radiation pattern. Such a tunable antenna is one of the critical elements for smart telecommunication systems in order to maintain various operating frequencies (to have more functionalities) with a single antenna unit (with a low volume and cost).

Reconfigurable filters also benefit from RF-MEMS capabilities (through low losses, notably). A 10% frequency tuning range is demonstrated in the X-band (Bouyge *et al*., 2012), which is attractive for the selection of various wireless communication standards with a single reconfigurable filter. Figure 16.12 illustrates such a filter, based on a split ring resonator in association with eight RF-MEMS switches.


METU 20KV — 100µm F1 L01 X55 38mm
METU 20KV 100µm F1 X350 37mm L01

16.11 RF-MEMS-based tunable dual-frequency antenna. (*Source*: Topalli *et al*. (2009).)


(a)
A
B
C
D
D
C
B
A
(b)
Off/up-state
MEMS beam-upper electrode
On/down-state
Insulating layer+ lower electrode
Cr
Au
Cr
V=0
V>V_P
Sapphire substrate

16.12 RF-MEMS-based tunable filter. (a) Filter topology and (b) RF-MEMS switches architecture. (*Source*: Bouyge *et al*. (2012).)

These demonstrations of the use of RF-MEMS switches and varactors to bring reconfigurability, tunability or smartness to microwave microsystems are but a few examples among many featuring a certain degree of agility in order to optimize the communication: standard one, operating frequency, operating power, etc. (Kim *et al*., 2001; Lu, 2003; Fukada 2004;

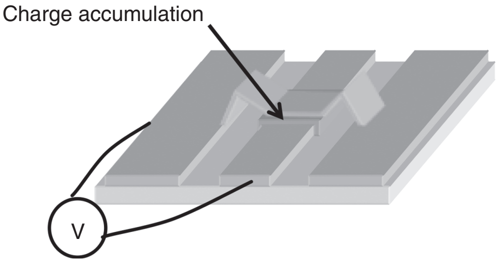
Hacker *et al*., 2004; de Graauw *et al*., 2006; Liu *et al*., 2011; Rantakari *et al*., 2011; Fouladi *et al*., 2012). From the perspective of industrial applications, additionally, another important issue is dealing with the reliability and the robustness of the different modules. They have to feature certain repairing and/or tuning functionalities in order to guarantee a high quality of service (QoS).

The RF-MEMS added value – that is to say, the high level of performance and large-scale integration in microsystems for smart communication systems – needs to be achieved without compromising reliability. This is crucial, especially if we consider commercial applications: what are the reliability issues with this technology and what are its limits, in terms of frequency, power and so on? The two next sections address the issues of reliability and the power capabilities of RF-MEMS devices.

16.4 RF-MEMS reliability

It is well known that the main cause of failure in capacitive RF-MEMS devices is due to dielectric charging, which causes electrostatic stiction (Van Spengen *et al*., 2002; Yuan *et al*., 2004; Mellé *et al*., 2005) of the mobile membrane over the coplanar line. The dielectric layer on top of the line then plays a central role in the reliability of capacitive RF-MEMS. Various dielectric materials are available and the required thickness is generally of the order of 0.2 μm, which allows the use of both polymer and mineral layers. The use of polymers may be possible but their nonconformal deposition and their low permittivity make mineral dielectrics better candidates. Dielectric layers such as SiO_2 or SiN obtained by PECVD are preferred because their low thermal budget is compatible with the metal of the transmission lines and also with possible active circuits located underneath. Their fabrication process, following deposition, traditionally involves reactive-ion etching (RIE) using a photoresist mask. To simplify understanding of the global process, the dielectric layer in Fig. 16.3b is realized on top of the transmission line, prior to the sacrificial layer being deposited. As this layer is present to prevent any direct contact between the transmitted signal line and the movable membrane, it may also be elaborated during the formation of the tunable part of the MEM element.

The physical phenomenon of failure is undesired stiction and operations of the membrane. This occurs because when voltage is applied to actuate the bridge or cantilever, charges accumulate in the dielectric. This phenomenon is illustrated in Fig. 16.13. To overcome this problem, several solutions have been investigated. One consists in replacing the dielectric layer with air. In order to avoid any contact between the membrane and the transmission line during actuation, dimples are placed on the movable membrane in front of the slots in the bottom metallic layer. This solution is efficient but limits the



16.13 Dielectric charging in capacitive RF MEMS.

performance of the switches in terms of the capacitance ratio and, consequently, possible applications.

Another investigation looked into doping the dielectric layer, in order to facilitate the evacuation of the charge during the application of the electrostatic force. This may be achieved by doping an SiN layer with Si, as patented by Goldsmith *et al*. (Ehmke *et al*., 2002). It may also be achieved by the incorporation of carbon nanotubes (CNTs) in the dielectric layer (Bordas *et al*., 2007). CNT doping translates into several orders of magnitude in improvement of the MEMS life span.

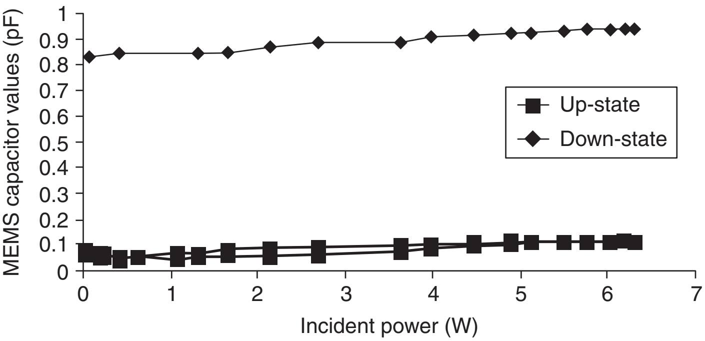
As far as the Ohmic-contact MEMS are concerned, a similar issue on the metal–metal contact has been studied and resolved. Compromises on performance are nevertheless noted, as good material for reliability features medium-resistive contact.

For both the Ohmic and the capacitive contacts, packaging is also a crucial issue (Wilkerson *et al*., 2001; Tilmans *et al*., 2003; Seki *et al*., 2004), as it impacts on the insertion losses and also on reliability since hermetic packaging may warrant a dried air supply to the MEMS. Moreover, the operation of the device inside the application and, especially, the level of power also figure in these options. This aspect is presented in the next section.

16.5 RF-MEMS power capability

As previously pointed out, the reliability of RF-MEMS needs to be studied with devices under the conditions of their future environment: temperature; gases, etc.; and RF-power. This section reports on the power behavior of RF-MEMS capacitive switches. We present investigations on the failure mechanism of capacitive RF-MEMS switches under high RF-power-monitoring (Pillans *et al*., 2002; Muldavin *et al*., 2003; Grenier *et al*., 2005; Peroulis *et al*., 2005).

To illustrate the degradation mechanism which appears under RF-power, two sets of experiments were performed on the device, as depicted in Fig. 16.4.

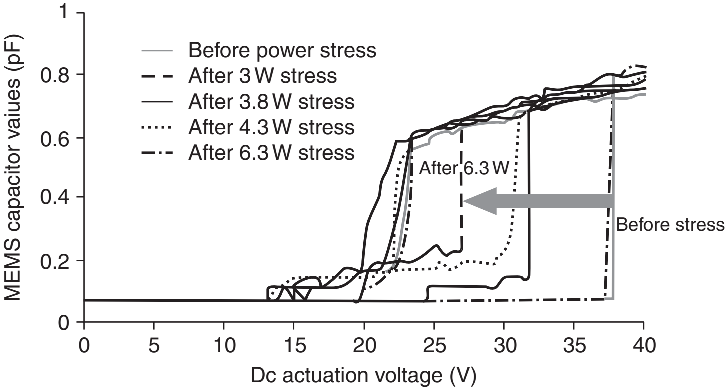


16.14 Measured capacitor values (up and down) *vs* RF-power stresses' intensity.

For the first set of experiments, the extracted up-state (Cup) and down-state (Cdown) capacitances of the RF-MEMS device after RF-power stresses were monitored. Before any stress was applied, the values of Cup and Cdown were extracted and corresponded to 65 fF and 0.84 pF, respectively. Figure 16.14 summarizes the experimental data with RF-power stress up to 6.3 W, the maximum power delivered by the setup. We report a 50 fF up-state capacitor rise which corresponds to only 0.12 dB insertion loss degradation at 10 GHz, which is negligible and proves that no self-actuation or microwave degradation occurs up to 6.3 W. Figure 16.14 also demonstrates no significant degradation of the down-state capacitor. On the contrary, an improvement of 12% is obtained, which translates into an isolation improvement of only 0.8 dB related to an enhancement of the contact quality by the RF-power.

A deeper insight of the power behavior of the switch is found with the second set of experiments. The rising RF-power stress from 0 to 6.3 W, when the bridge is in the down state, was induced and the electromechanical behaviors of the RF-MEMS were monitored (we have already demonstrated that no degradation of microwave performance appears). Figure 16.15 presents the measurements of the microwave transmission (parameter S_{21}) of the RF-MEMS (typically, at 10 GHz) for DC actuation voltage ranging from 0 to 40 V and down to 0 V. From this measurement, the behavior vs. the RF-power stress of the insertion losses of the MEMS switch (S_{21} for 0 V_{DC}), the isolation (S_{21} for 40 V_{DC}) and the pull up and down threshold voltages can be monitored.

The observed behavior is associated with the fact that, up to 3 W, no degradation of the device may be noted. Degradation occurs between 3 and 3.8 W. For power stress greater than 4 W, a magnification of the degradation mechanism reflected by a different electromechanical behavior is shown. The degradation mechanism signature corresponds to a decrease of the



16.15 Microwave behavior at 10 GHz *vs* DC control voltage for different power stress levels.

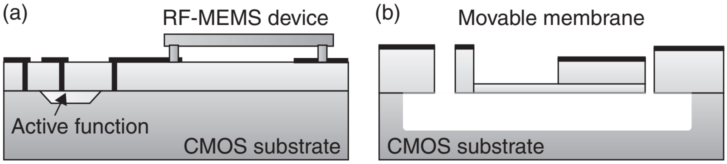
pull-in voltage, which is related to a modification of the overall stiffness. The combination of these experiments and the observed results suggest that a mechanical degradation of the movable membrane occurs at high RF-power, consequently to high current density above the electromigration threshold, as will be discussed in the next paragraph.

In order to understand the causes of such failure on the electromechanical behavior, electromagnetic simulations with SONNET software permit the prediction of the RF current density at 10 GHz. For 3 and 3.8 W of injected power, current density of 112 and 124 GA/m^2 for up and down bridge states were simulated. These results are very close to the electromigration threshold values reported in the literature for gold metallization (around 140 GA/m^2 at 80°C). This accord between extracted failure current density and the electromigration threshold strongly hypothesizes that the main failure phenomenon occurs in the mobile part of the RF-MEMS and translates into degradation in electromechanical performance.

16.6 Co-integration of RF-MEMS-based circuits with integrated circuits (IC)

We have already discussed the weak points of RF-MEMS technologies, into which considerable research is being undertaken. These drawbacks do not have to hide and prohibit research on one of the most attractive features of RF-MEMS devices: their integration with both active and passive functions with no compromise on microwave performance and system miniaturization.

One key issue of RF-MEMS technology is, then, to be compatible with CMOS platforms. Such compatability opens the door to the



16.16 Placement of RF-MEMS with IC. (a) RF-MEMS above-IC and (b) RF-MEMS in-IC.

co-integration of passive and active functions with RF-MEMS, which are the key to bringing high-level performances in tunability or reconfigurability to the systems. One critical element of which to be aware for such co-integration consists in the tolerable thermal budget. As far as CMOS technology is concerned, this limit is traditionally considered to be around 400°C. Above this temperature, the circuits begin to degrade (Grenier *et al.*, 2004).

The second key factor in the integration of RF-MEMS passive components with ICs is related to cost. IC substrate is particularly expensive because of the demanding process steps and photolithographic masks required for the elaboration of transistors and integrated passive components. Consequently, if the RF-MEMS passive elements require a large surface area, their monolithic integration on IC substrate would become cost-ineffective. The choice between monolithic and hybrid integrations is thus, essentially, driven by the surface consumption of the different parts to be co-integrated. A trade-off has to be evaluated between the benefits to performance and the global cost. In the case of hybrid integration, different techniques exist to interconnect active circuits with MEMS technologies: wire bonding and flip-chip implementations. In this context, RF-MEMS die varactors were co-integrated (using the flip-chip method) on a laminate board with a BiCMOS class-F power amplifier. This provides an up to 10% improvement in efficiency for a dual-band 0.9/1.8 GHz operation (Graauw, 2006).

As far as the monolithic integration of RF-MEMS is concerned, it is traditionally realized through post processing steps, after the realization of the ICs, as indicated in Fig. 16.16a. For example, Nguyen (2008) demonstrated vibrating disk mechanical resonators realized on a CMOS substrate to elaborate integrated micromechanical radio front-ends.

More recently, the new opportunities offered by thick dielectric and metallic layers with IC technologies has opened new horizons for integrated MEMS with ICs (Fedder *et al.*, 2008), as indicated in Fig. 16.16b. An excellent tuning ratio of digital capacitors of up to 60:1 has been demonstrated with the employment of the top dielectric and metallic layer available on ICs (Reinke *et al.*, 2011).

Monolithic integration attracts increasing attention since the interconnections do not degrade the performances available from RF-MEMS devices. This enables the tuning or reconfiguration of systems with no compromise in performance, which is essential if the target is to improve the power efficiency of an amplifier, for example. In this context, an RF-MEMS SPDT integrated on a gallium arsenide (GaAs) platform together with a low-noise amplifier permit the provision of up to 1 dB noise figure improvement for a wide K-band operation (Malmqvist *et al.*, 2012). Moreover, this monolithic integration has been reached with no surface counterpart, as the final MMIC surface is only 3 mm^2. Ulusoy *et al.* (2012) demonstrated a similar benefit on a BiCMOS platform. A 60–77 GHz switchable low-noise amplifier integrated with RF-MEMS switches has been realized with a 0.25 μm SiGe-C BiCMOS technology. Even if the two RF-MEMS devices occupy a large surface compared with the low-noise amplifier (LNA) core, the circuit occupies only 0.8 mm^2, which validates the readiness of RF-MEMS devices to be integrated monolithically with active functions.

16.7 Conclusion

RF-MEMS technologies have been successfully introduced since the turn of the millennium in order to develop smart microsystems exhibiting a high level of integration, new functionalities such as reconfigurability (the ability to switch between different standards) or to self-repair, and high-level electrical performance with up to millimeter wave frequencies (Rebeiz and Muldavin, 2001).

Despite the fact that numerous demonstrations of the very promising capabilities offered by RF-MEMS have been shown (Rebeiz, 2003), there are still key issues that are under investigation, such as reliability, packaging and power-handling. Regarding reliability, considerable progress has been made in the understanding of the dielectric charging phenomenon and some solutions have been proposed to reduce this effect and to enhance the life span of the device (Bordas *et al.*, 2007). Concerning packaging, there is a great deal of ongoing research into finding a solution that would be ‘MEMS’compatible (low thermal and stress budgets) and that features a certain level of hermeticity (Wilkeson *et al.*, 2001; Tilmans *et al.*, 2003; Seki, 2004). The issue regarding hermeticity – probably the most important issue, as it essentially concerns microwave behavior – deals with the power-handling capabilities of MEMS devices.

Beyond these reliability issues, the major capabilities of RF-MEMS technologies should be pointed out: the monolithic devices on GaAs and

BiCMOS platforms, and the extensive integration of RF-MEMS with passive circuits with no compromise in performance. The first capability opens the door to the ultimate miniaturization of smart systems; the second targets exceptionally high-level performance for reconfigurable or tunable functions.

For both, RF-MEMS technology has already been demonstrated as a royal route to reconfigurability, tunability and (more generally) smartness in RF, microwave and millimeter wave systems.

16.8 References

Bordas, C., Grenier, K., Dubuc, D., Flahaut, E., Pacchini, S., Paillard, M. and Cazaux, J. L. (2007), 'Carbon nanotube based dielectric for enhanced RF MEMS reliability', *IEEE International Microwave Symposium*, 3–8 June 2007, Honolulu, USA.

Bouyge, D., Crunteanu, A., Durán-Sindreu, M., Pothier, A., Blondy, P., Bonachel, J., Orlianges, J.-C. and F. Martin (2012), 'Reconfigurable split rings based on MEMS switches and their application to tunable filters', *Journal of Optics,* **14**, 11, 114001.

Dubuc, D., Rabbia, L., Grenier, K., Pons, P., Vendier, O., Graffeuil, J. and Plana, R. (2003), 'Original MEMS-based single pole double throw topology for millimeter wave space communications', *33rd European Microwave Conference*, 6–10 October 2003, Munich, Germany, 3, 979–982.

Ducarouge, B., Dubuc, D., Melle, S., Grenier, K., Bary, L., Pons, P. and Plana, R. (2004), 'Efficient design methodology of polymer based RF MEMS switches', *5th IEEE Topical Meeting on Silicon Monolithic Integrated Circuits in RF systems (SiRF)*, 8–10 September 2004, Atlanta, 298–301.

Ehmke, J. C., Goldsmith, C. L., Yao, Z. J. and Eshelman, S. M. (2002), 'Method and apparatus for switching high frequency signals', Patent No*: US 6,391,675 B1.*

Fedder, G. K., Howe, R. T., King, Liu T.-J. and Quevy, E. P. (2008), 'Technologies for cofabricating MEMS and electronics', *Proceedings of the IEEE*, **96**, 2, 306–322.

Fouladi, S., Domingue, F. and Mansour, R. (2012), 'CMOS-MEMS tuning and impedance matching circuits for reconfigurable RF front-ends', *Microwave Symposium Digest (MTT), 2012 IEEE MTT-S International*, 17–22 June 2012, Montreal, Quebec, 1–3.

Fukada, A. (2004), 'Novel 900 MHz/1.9 GHz dual-mode power amplifier employing MEMS switches for optimum matching', *IEEE Microwave and Wireless Components Letters,* **14**, 3, 121–123.

Goldsmith, C., Lin, T.–H., Powers, B., Wu, W.–R. and Norvell, B. (1995), 'Micromechanical membrane switches for microwave applications', *Technical Digest, IEEE MicrowaveTheory and Techniques Symposium*, 91–94.

Goldsmith, C., Randall, J., Eshelman, S., Lin, T.–H., Denniston, D., Chen, S. and Norvell, B. (1996), 'Characteristics of micromachined switches at microwave frequencies', *Technical Digest, IEEE Microwave Theory and Techniques Symposium,* 1141.

De Graauw, A. J. M., Steeneken, P. G., Chanlo, C., Dijkhuis, J., Pramm, S., Van Bezooijen, A., Ten Dolle, H. K. J., Van Straten, F. and Lok, P. (2006), 'MEMS-based reconfigurable multi-band BiCMOS power amplifier', *Bipolar/BiCMOS Circuits and Technology Meeting*, 8–10 October 2006, Maastrict, The Netherlands, 1–4.

Grenier, K., Dubuc, D., Mazenq, L., Busquère, J. P., Ducarouge, B., Bouchriha, F., Rennane, M., Lubecke, V., Pons, P., Plana, J. R. and Ancey, L. P. (2004), 'Polymer based technologies for microwave and millimeter wave applications', *50th IEEE International Electron Devices Meeting*, 13–15 December 2004, Washington, USA.

Grenier, K., Dubuc, D., Ducarouge, B., Conedera, V., Bourrier, D., Ongareau, E., Derderian, P. and Plana, R. (2005), 'High power handling RF MEMS design and technology', *18th IEEE International Conference on Micro Electro Mechanical Systems (MEMS '2005)*, 30 January–3 February 2005, Miami, USA, 155–158.

Hacker, J. B., Kim, M., Mihailovich, R. E. and DeNatale, J. F. (2004), 'Monolithic GaAs PHEMT MMICs integrated with RF MEMS switches', *IEEE 2004 CSIC Digest,* 24–27 October 2004, Monterey, USA 229–232.

Halg, B. (1990), 'On a micro-electro-mechanical nonvolatile memory cell,' *IEEE Transactions on Electron Devices*, 37, 2230–2236.

Kim, M., Hacker, J. B., Mihailovich, R. E. and DeNatale, J. F. (2001), 'A monolithic MEMS switched dual-path power amplifier', *IEEE Microwave and Wireless Components Letters,* **11**, 7, 285–286.

Larson, L. E., Hackett, R. H., Melendes, M. A. and Lohr, R. F. (1991), 'Micromachined microwave actuator (MIMAC) technology – a new tuning approach for microwave integrated circuits', *Microwave and Millimeter-Wave Monolithic Circuits Symposium Digest*, 10–11 June 1991, Boston, USA, 27–30.

Liu, R., Schreurs, D., De Raedt, W., Vanaverbeke, F. and Mertens, R. (2011), 'RF-MEMS based tri-band GaN power amplifier', *Electronics Letters*, **47**, 13, 762–763.

Lu, Y. (2003), 'A MEMS reconfigurable matching network for class AB amplifier', *IEEE Microwave and Wireless Components Letters,* **13**, 10, 437–439.

Malmqvist, R., Samuelsson, C., Gustafsson, A., Maher, H., Vaha-Heikkila, T. and Baggen, R. (2012), 'A K-band single-chip reconfigurable/multi-functional RF-MEMS switched dual-LNA MMIC', *Microwave Symposium Digest (MTT), 2012 IEEE MTT-S International*, 17–22 June 2012, Montreal, Canada, 1–3.

Mellé, S., De Conto, D., Dubuc, D., Grenier, K., Vendier, O., Muraro, J. L., Cazaux, J. L. and Plana, R. (2005), 'Reliability modeling of capacitive RF-MEMS', *IEEE Transactions on Microwave Theory and Techniques,* **53**, 11, 3482–3488.

Muldavin, J. B. and Rebeiz, G. M. (1999), '30 GHz tuned MEMS switches', *Technical Digest, IEEE Microwave Theory and Techniques Symposium*, 13–19 June 1999, Anaheim, USA, 1511–1514.

Muldavin, J., Boisvert, R., Bozler, C., Rabe, S. and Keast, C. (2003), 'Power handling and linearity of MEM capacitive series switches', *IEEE MTT-S International Microwave Symposium Digest*, 8–13 June 2003, Philadelphia, USA, 1915–1918.

Nguyen, C. T. C. (2008), 'Integrated micromechanical radio front-ends', *IEEE International Symposium on VLSI Technology, Systems and Applications*, 21–23 April 2008, Hsinchu, China, 3–4.

Oberhammer, J., Tang, M., Liu, A. Q. and Stemme, G. (2006), 'Mechanically tri-stable in-line single-pole-double-throw all-metal switch', *IEEE MEMS Conference*, 22–26 January 2006, Istanbul, Turkey, 898–901.

Pacheco, S., Nguyen, C. T. and Katehi, L. P. B. (1998), 'Micromachined electrostatic K-band switches', *Technical Digest IEEE MTT-S International Microwave Symposium*, 7–12 June 1998, Baltimore, USA, 1569–1572.

Pacheco, S. P., Peroulis, D. and Katehi, L. P. B. (2001), 'MEMS single-pole double-throw (SPDT) X and K-band switching circuits', *IEEE MTT-S*, 20–24 May 2001, Phoenix, USA.

Palei, W., Liu, A. Q., Yu, A. B., Alphones, A. and Lee, Y. H. (2005), 'Optimization of design and fabrication for micromachined true time delay (TTD) phase shifters', *Sensors and Actuators A: Physical,* **119**, 2446–454.

Peroulis, D., Pacheco, S. P. and Katehi, L. P. B. (2005), 'High-power MEMS varactors and impedance tuners for millimeter-wave applications', *IEEE Transactions on Microwave Theory and Techniques*, **53**, 11, 3672–3678.

Petersen, K. E. (1979), 'Micromechanical membrane switches on silicon', *IBM Journal of Research Development*, **23**, 4, 376–385.

Pillans, B., Kleber, J., Goldsmith, C. and Eberly, M. (2002), 'RF power handling of capacitive RF devices', *IEEE MTT-S International Microwave Symposium Digest*, 2–7 June 2002, Seattle, USA, 329–332.

Qiao, D., Molfino, R., Lardizabal, S. M., Pillans, B., Asbeck, P. M. and Jerinic, G. (2005), 'An intelligently controlled RF-power amplifier with a reconfigurable MEMS-varactor tuner', *IEEE Transactions on Microwave Theory and Techniques*, **53**, 3, 2, 1089–1095.

Rantakari, P., Malmqvist, R., Samuelsson, C., Leblanc, R., Smith, D., Jonsson, R., Simon, W., Saijets, J., Baggen, R. and Vähä-Heikkilä, T. (2011), 'Wide-band RF MEMS switches and switching circuits using a gallium arsenide monolithic microwave integrated circuits foundry process technology', *IET Microwaves, Antennas & Propagation,* **5**, 8, 948–955.

Rebeiz, G. M. (2003), *RF MEMS: Theory, Design and Technology*, Hoboken USA: Wiley.

Rebeiz, G. M. and Muldavin, J. B. (2001), 'RF MEMS switches and switch circuits', *IEEE Microwave Magazine*, **2**, 4, 59–71.

Reinke, J., Fedder, G. K. and Mukherjee, T. (2011), 'CMOS-MEMS 3-bit digital capacitors with tuning ratios greater than 60:1', *IEEE T-MTT*, **59**, 5, 1238–1248.

Seki, T. (2004),' Recent progress in packaging of RF MEMS', *IEEE Compound Semiconductor Integrated Circuit Symposium*, 24–27 October 2004, Monterey, USA, 233–236.

Shen, Q. and Baker, N. S. (2005), 'A reconfigurable RF MEMS based double slug impedance tuner', *European Microwave Conference*, 4–6 October 2005, Paris, France, 537–540.

Sovero, E. A., Mihailovich R., Deakin, D. S., Higgins, J. A., Yao, J. J., DeNatale, J. F. and Hong, J. H. (1999), 'Monolithic GaAs PHEMT MMICs integrated with high performance MEMS microrelays', *Proceedings of IMOC '99*, 9–12 August 1999, Rio de Janeiro, Brazil.

Tang, M., Yu, A. B., Liu, A. Q., Agarwal, A., Aditya, S. and Liu, Z. S. (2005), 'High isolation X-band MEMS capacitive switches', *Sensors and Actuators A: Physical,* **120**, 1, 241–248.

Tilmans, H. A. C., De Raedt, W. and Beyne, E. (2003), 'MEMS for wireless communications: from RF MEMS components to RF-MEMS-SIP', *Journal of Micromechanics and Microengineering*, **13**, 4, 139–163.

Topalli, K., Erdil, E., Aydin, Civi O., Demir, S., Koc, S. and Akin, T. (2009), 'Tunable dual-frequency RF MEMS rectangular slot ring antenna', *Sensors and Actuators A: Physical*, **156**, 2, 373–380.

Ulusoy, A. C., Kaynak, M., Purtova, T., Tillack, B. and Schumacher, H. (2012), 'A 60 to 77 GHz switchable LNA in an RF-MEMS embedded BiCMOS technology', *IEEE Microwave and Wireless Components Letters,* **22**, 8, 430–432.

Vähä-Heikkilä, T. and Rebeiz, G. M. (2004), 'A 4-18-GHz reconfigurable RF MEMS matching network for power amplifier applications', *International Journal of RF and Microwave Computer-Aided Engineering,* **14**, 4, 356–372.

Van Spengen, W. M., Puers, R., Mertens, R. and De Wolf, I. (2002), 'Experimental characterization of stiction due to charging in RF MEMS', *International Electron Devices Meeting 2002*, 8–11 December 2002, San Francisco, USA, 901–904.

Wilkerson, P., Kranz, M., Przekwas, A. and Hudson, T. (2001), 'Flip chip hermetic packaging of RF MEMS', *Microelectromechanical Systems Conference*, 20–24 January, Las Vegas, USA, 91–94.

Yao, Z. J., Chen, S., Eshelman, S., Denniston, D. and Goldsmith, C. (1999), 'Micromachined low-loss microwave switches', *IEEE Journal of MEMS,* **8**, 129–134.

Yuan, X., Cherepko, S., Hwang, J., Goldsmith, C. L., Nordquist, C. and Dyck, C. (2004), 'Initial observation and analysis of dielectric-charging effects on RF MEMS capacitive switches', *2004 IEEE MTT-S Int. Microwave Symposium Digest*, 6–11 June 2004, Fort Worth, USA, 1943–46.

17

Smart acoustic sensor array (SASA) system for real-time sound processing applications

M. TURQUETI, E. ORUKLU and J. SANIIE,
Illinois Institute of Technology, USA

DOI: 10.1533/9780857099297.2.492

Abstract: This chapter describes the design and implementation of a smart acoustic sensor array (SASA) system that consists of a 52-microphone micro-electromechanical systems array embedded in a field-programmable gate array platform with real-time data acquisition, signal processing and network communication capabilities. The SASA system is evaluated using several case studies in order to demonstrate the versatility and scalability of the sensor array platform for real-time sound analysis applications. These case studies include sound mapping and source localization.

Key words: acoustic MEMS array, sound localization, field-programmable gate arrays, smart sensors.

17.1 Introduction

Designing a smart acoustic sensor array (SASA) system presents significant challenges due to the number of sensors, the processing speed and the complexity of the targeted applications. This chapter describes the design and implementation of a 52-microphone micro-electromechanical systems (MEMS) array, embedded in an field-programmable gate array (FPGA) platform with real-time processing capabilities. In this type of system, characteristics such as speed, scalability and real-time signal processing are paramount and, furthermore, highly advanced data acquisition and processing modules are necessary. Sound processing applications such as sound source mapping, source separation and localization have several important characteristics that the system must meet – such as array spatial resolution, low reverberation and real-time data acquisition and processing.

The acoustic MEMS array (AMA) system presented in this study is designed to meet these challenges. The design takes into account mechanical factors such as array geometry, sensor disposition and microphone reverberation. In addition, the design of the electronics system addresses electronic noise, power decoupling, cross-talk and connectivity.

The system integrates the acoustic array with real-time data acquisition, signal processing and network communication capabilities. FPGA-based smart sensor nodes are examined in several case studies, demonstrating the versatility and scalability of the sensor array platform for real-time applications. These case studies include sound mapping and source localization.

Microphone arrays are capable of providing spatial information for incoming acoustic waves. Arrays can capture key information that would be impossible to acquire with single microphones; however, this additional information comes at a price in terms of the increased complexity of the system. The design of microphone arrays faces several challenges – such as number of detectors, array geometry, reverberation, interference issues and signal processing (Buck *et al.*, 2006). These factors are crucial to the construction of a reliable and effective microphone array system.

Microphone arrays present additional challenges for signal processing methods since large numbers of detecting elements and sensors generate large amounts of data to be processed. Furthermore, applications such as sound tracking and sound source localization require complex algorithms in order to process the raw data appropriately (Benesty *et al.*, 2008). Challenging environments such as multiple sound sources moving with background noise are especially difficult to deal with when real-time processing is required. It is also important for such a system to be easily scalable, as demonstrated by Weinstein *et al.* (2004) and Benesty *et al.* (2008). The performance of a microphone array typically increases linearly with the size of the array.

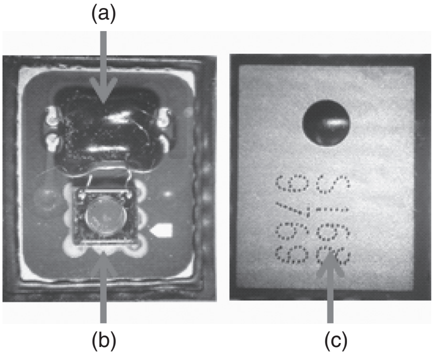
The system presented in this project is designed to provide an acoustic data acquisition system that is flexible and expandable, with powerful real-time signal processing capability in order to be interfaced with a wide variety of sensor arrays. The data acquisition and processing architecture presented in this work is called the 'compact and programmable daTa acquisition node' (CAPTAN) (Turqueti *et al.*, 2008) and the acoustic array embedded on the system is known as an AMA. The combination of AMA and CAPTAN is known as a SASA system.

The CAPTAN architecture is a distributed data acquisition and processing system that can be employed in a number of different applications ranging from a single sensor interface to multi-sensor arrays data acquisition and to high-performance parallel computing (Rivera *et al.*, 2008). This architecture has the unique features of being highly expandable, interchangeable and adaptable and with a high computation power inherent in its design. The AMA array was designed to conform to and to take advantage of the CAPTAN architecture. It is an acoustic array that employs sound or ultrasound sensors distributed in two dimensions.

17.2 MEMS microphones

MEMS microphones have being introduced onto the market in large numbers by the advent of mobile communication devices. Mobile phones require inexpensive, robust and reliable microphones; MEMS microphones provide a perfect match. Since the introduction of the MEMS microphone, its performance has improved vastly. It now has comparable or better performance than conventional magnetic or piezo microphones, especially with regard to frequency response and the signal-to-noise ratio.

There are many possible implementations of MEMS microphones; the implementation utilized in this work is based on capacitive transduction. Capacitive MEMS microphone operation is based on the change of the capacitance of the sensor element within the microphone when acoustic waves interact with the microphone. This is usually achieved by placing a metalized membrane a few microns from a metalized back plate, where the air gap is the dielectric. The membrane distance to the backplane fluctuates based on air pressure changes caused by incoming acoustic waves. System operation is analogous to a parallel plates capacitor. The MEMS microphone utilized on this project also employs an acoustic chamber to increase the acoustic pressure on the membrane and to improve its gain. In order to translate the small capacitive oscillations to electrical signals, a pre-amplifier is embedded in the same package. Usually, the pre-amplifier is built using standard complementary metal-oxide-semiconductor (CMOS) processes, while the membrane is deposited in additional steps. Frequently, these additional steps involve the sputtering of several metal layers, photoresist processes and polyimide. For the speaker microphone SPM 208 microphone (Knowles, 2006), the sensor element and the pre-amplifier are built in two different wafers, comprising two chips that are electrically connected by wire bonds. The microphone membrane was built on top of a silicon substrate (see Fig. 17.1).



17.1 MEMS microphone utilized in this work. (a) Pre-amplifier, (b) acoustic transducer and (c) aluminum cover forming resonant cavity.

17.3 Fundamentals of acoustic sensor arrays and applications

There is currently a significant amount of research and numerous applications which use sound or ultrasound for communications, detection and analysis. Some of the research topics are multi-party telecommunications; hands-free acoustic human-machine interfaces; computer games; dictation systems; hearing-aids; medical diagnostics; the structural failure analysis of buildings or bridges, and the mechanical failure analysis of machines such as vehicles or aircraft; and robotic vision, navigation and automation (Llata *et al*., 2002; Nishitani *et al*., 2005; Alghassi, 2008; Kunin *et al*., 2010, 2011; Eckert *et al*., 2011; Kim and Kim, 2011). In practice, there are numerous issues encountered in the real-world environment which make the realistic application of this theory significantly more difficult (Chen, 2002; Rabinkin *et al*., 1996; Brandstein and Ward, 2001). These issues include ambient sound and electrical noise; the presence of wideband nonstationary source signals; the presence of reverberation echoes; high-frequency sound sources which require higher speed systems; and the fluctuation of the ambient temperature and humidity, which affect the speed at which sound waves propagate.

The geometry of the array can play an important part in the formulation of the processing algorithms (Benesty *et al*., 2008). Different applications require different geometries in order to achieve optimum performance. In applications such as source tracking, the array geometry is very important in determining the performance of the system (Nakadai *et al*., 2002). Regular geometries, where sensors are evenly spaced, are preferred in order to simplify development of the algorithm. Linear arrays are usually applied in medical ultrasonography, planar arrays are often used in sound source localization and three dimensional spherical arrays are most frequently used in sophisticated SONAR applications. In other applications, such as source separation, the geometry of the array is not as important as the transducer characteristics – such as dynamic range and the transducer aperture. The size of an array is usually determined by the frequency at which the array will operate and the kind of spatial resolution required by the application using the acoustic array (Weinstein *et al*., 2004).

A data acquisition system is necessary to condition, process, store and display the signals received by the array. Data acquisition for an acoustic array can be very challenging, but most of the issues are co-related with the types of sensor, number of sensors and array geometry. Also, the application plays an important role in defining the needs of the array for signal processing or real-time operation requirements.

Beamforming is a technique used in transducer arrays for establishing the signal beam field and directivity for transmission or reception. The spatial directivity is achieved by the use of interference patterns to change the angular directionality of the array. When used for the transmission of signals, the transmitting signal will be steered – this is where the amplitude and phase of its individual elements are controlled to act in unison through patterns of constructive and destructive interference. The resulting wavefront will contentrate the energy of the array in the desired direction. There are two categories of beamforming; static and adaptive (Hodgkiss, 1980; Campbell, 1999). Static beamforming involves the use of a fixed set of parameters for the transducer array. Adaptive beamforming, on other hand, can adapt the parameters of the array in accordance with changes in the application environment. Adaptive beamforming, although computationally demanding, can offer better noise rejection than static beamforming.

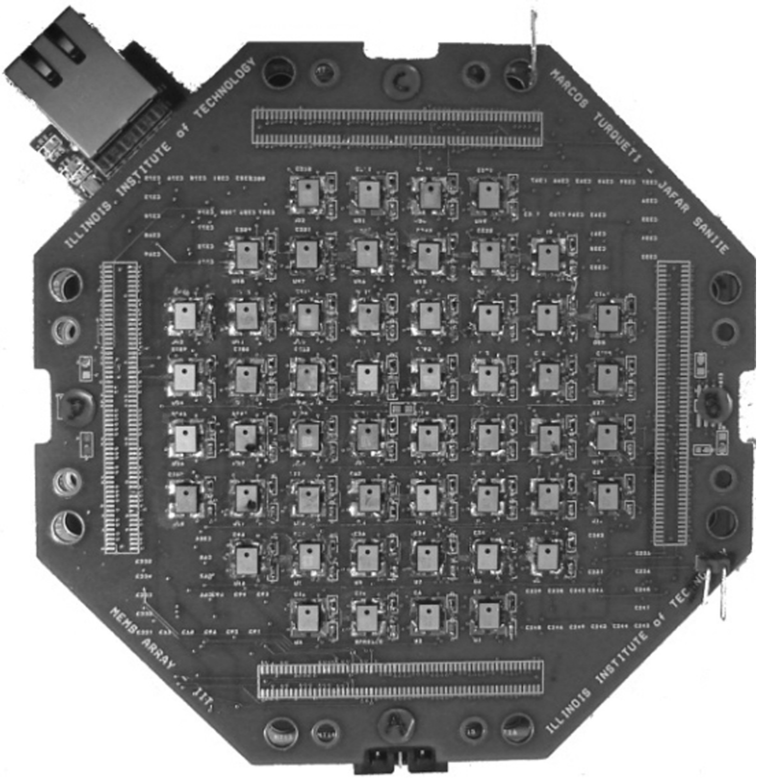
17.4 Design and implementation of a smart acoustic MEMS array (AMA)

The design process for the AMA involves selecting the optimum topological distribution of sensors, number of microphones, inter-microphone spacing and microphone type. The presented acoustic array was designed to operate within sonic frequencies ranging from 100 Hz to 17 kHz, while maintaining a flat response within 10 dB between 100 Hz and 10 kHz. Also, for beamforming, the spatial resolution should allow the sampling of frequencies of 17 kHz. In order to avoid spatial aliasing and to obtain a good acoustic aperture, the inter-microphone distance was determined to be 10.0 mm from center to center. This spacing makes it possible to obtain relevant phase information of incoming acoustic sound waves, increasing the array sensitivity and allowing spatial sampling of frequencies up to 17 kHz without aliasing. The inter-microphone array spacing of 10.0 mm was calculated by finding the wavelength of the desired 17 kHz upper limit. Utilizing Equation [17.1]:

$$c = \lambda f_u \qquad [17.1]$$

where c is the speed of sound (340 m/s), f_u is the upper limit frequency and λ is the upper limit wavelength. Solving Equation [17.1], λ was calculated to be 2.00 cm. The inter-microphone is calculated to satisfy the Nyquist–Shannon sampling theorem, and is therefore equal to the upper wavelength divided by two, which results in the array inter-microphone spacing being 1.00 cm (i.e., 10.0 mm).

The number of microphones is a compromise between the size of the board, the electronics needed to deal with the massive amount of data and the need to achieve a signal-to-noise ratio of at least 20 dB. The AMA board (see Fig. 17.2) consists of 52 MEMS microphones distributed in an


ILLINOIS INSTITUTE of TECHNOLOGY
MARCOS TURQUETI - JAFAR SANIIE
MEMS ARRAY IIT
ILLINOIS INSTITUTE of TEC

17.2 Top view of the CAPTAN readout system. with AMA board hosting the 52 MEMS microphones. This package is named smart acoustic sensor array (SASA).

octagonal layout of 8 columns and 8 rows with 2 central transmitter elements. The central elements can be used for calibration purposes or sonar applications. When used as a calibration element, the central element emits a set of mono-component frequencies. These signals are then captured by the microphones for calibrating the response of the array, taking into account the spatial distribution of microphones. When the central element is used for active sonar applications, a series of pre-programmed pulses are emitted and then, when the emitted waves encounter obstacles, they bounce back to the array, giving information about the distance and position of obstacles.

The MEMS microphones are the fundamental components of the array, and their design and fabrication must offer high sensitivity and low reverberation noise. The overall sensitivity of the array increases monotonically with the number of sensors. The MEMS microphone chosen for this array (the SPM208) has a sensitivity of 1 V/Pa at 1 kHz (Knowles, 2006). Figure 17.1 shows the selected MEMS microphone. These microphones are omni directional and, when combined into an array, they provide a highly versatile acoustic aperture. The selected MEMS microphone frequency response is essentially flat from 1 to 8 kHz, and has a low frequency limit of 100 Hz and a high-frequency limit of 25 kHz. The microphones are glued with silver epoxy to the copper pads in the board in order to shield the microphone response against acoustic noise and reverberation.

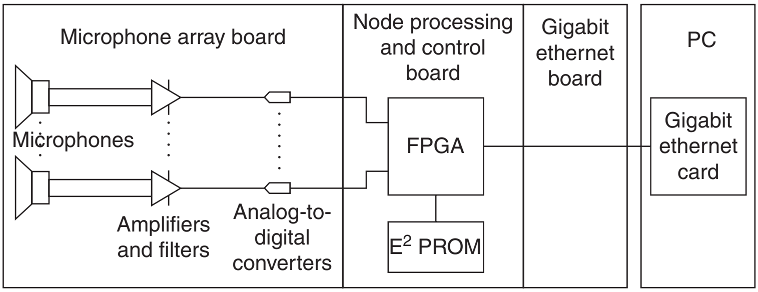
The number of microphones determines the array gain. The array gain (G) can be calculated with Equation [17.2]:

$$G = \frac{\mathrm{SNR_A}}{\mathrm{SNR_M}} \qquad [17.2]$$

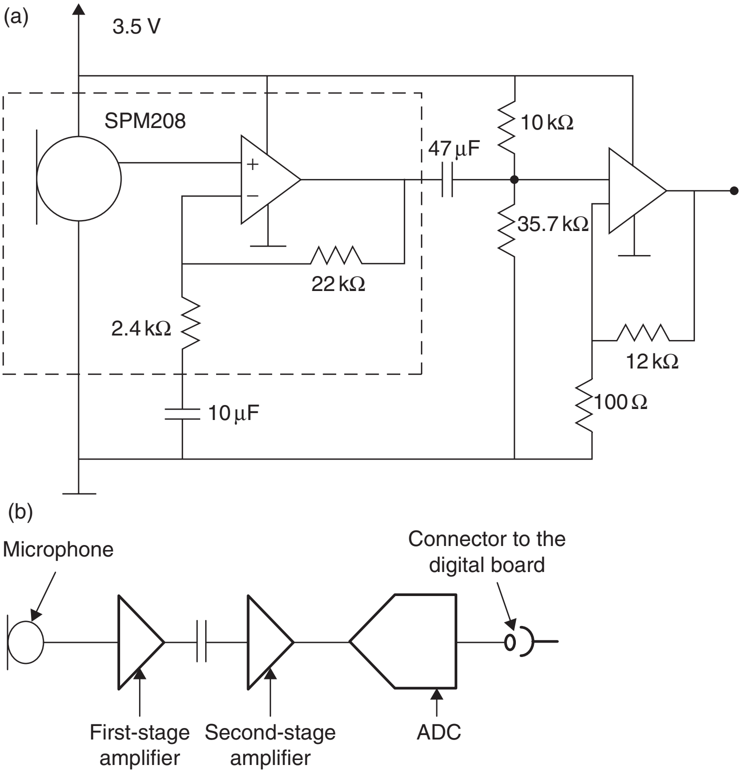
where $\mathrm{SNR_M}$ is the signal-to-noise ratio of the microphone and $\mathrm{SNR_A}$ the signal-to-noise ratio of the array. $\mathrm{SNR_A}$ for this acoustic array – and for planar acoustic arrays, in general – increases linearly with the increasing number of microphones. For beamforming applications, it is desirable to make the array gain (G) at least 2. For the proposed geometry, the increase of the $\mathrm{SNR_A}$ is linear and was experimentally determined to be 1.3 dB per added microphone. Therefore, to determine the $\mathrm{SNR_A}$ for this array we multiply the number of microphones minus one by 1.3 dB and add the $\mathrm{SNR_M}$. The microphones have an $\mathrm{SNR_M}$ of 59 dB at 1 kHz which results in an array gain of 2.12 at 1 kHz.

In support of the microphones, the AMA board also provides analog-to-digital convertors that interface with the CAPTAN board through the four board-to-board vertical bus connectors (Turqueti *et al.*, 2008). The readout system is based on the CAPTAN architecture and its implementation is presented in the next section. Physically, the system can be separated into three parts: the AMA, node processing and control board (NPCB) and the Gigabit Ethernet Link (GEL) board. The NPCB is the backbone board that contains the FPGA which contains the system's configware. The GEL board controls Ethernet communications. The NPCB and GEL boards are part of the CAPTAN system. The AMA board is the hardware which contains the microphones, amplifiers and analog-to-digital converters (ADCs). Figure 17.3 shows the three hardware components that make up the SASA system.

The board containing the MEMS array is responsible for data acquisition. Each of the 52 MEMS microphones is equipped with a front-end signal conditioning circuit. This circuit is presented in Fig. 17.4a and 17.4b. It has the objective of amplifying and filtering the microphone signal. The first stage of the amplifier gives a 20 dB gain and is embedded within the microphone

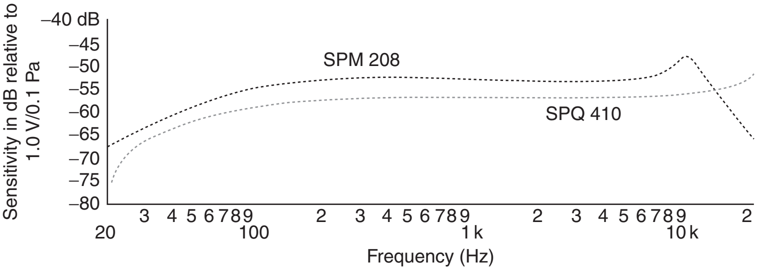

Microphone array board
Microphones
Amplifiers and filters
Analog-to-digital converters
Node processing and control board
FPGA
E^2 PROM
Gigabit ethernet board
PC
Gigabit ethernet card

17.3 Functional block diagram of the SASA system.


(a)
3.5 V
SPM208
+
−
47 µF
10 kΩ
35.7 kΩ
22 kΩ
2.4 kΩ
10 µF
12 kΩ
100 Ω
(b)
Microphone
Connector to the digital board
First-stage amplifier
Second-stage amplifier
ADC

17.4 (a) Front-end electronics of the AMA array and (b) two-stage amplifier circuit for each microphone.

case, as indicated by the dashed area in Fig. 17.4b. The second stage of the amplifier gives a further 20 dB of gain where the gain can be adjusted by the feedback resistor. The amplifiers also provide a second-order high-pass filter with a low cut-off frequency of 400 Hz.



17.5 Comparison of the sensitivity of two acoustic MEMS microphones. The line on the top shows the frequency response *vs* the sensitivity chart for the microphone utilized in this work (SPM 208); the line on the bottom shows a competing microphone (SPQ 410) that was discarded due to its lower sensitivity.

This board supports two different commercially available MEMs microphones from Akustica: the SPM208, with a dynamic range varying from 100 Hz to 12 kHz; and the SPM204, with a dynamic range varying from 10 to 65 kHz. The microphones can be intermixed on the array or, alternatively, the array could be fabricated using one type of microphone only. The mix of the two different sensors allows sound analysis and ultrasound ranging applications. The frequency response of the SPN208 microphone is shown in Fig. 17.5.

After the analog signal has been conditioned, it is processed by the channel ADC. Each channel has its own ADC: an ADC121S101 (Texas Instruments, 2010). This is a serial 12-bit A to D converter with a maximum sampling rate of 1 M samples/s. The system presented in this work was set to work at 36 K samples/s but the configware can be simply adjusted to any sampling rate that is desired up to 1 M samples/s. After the ADC digitizes the signal, it generates a serial bit stream at 432 K bits/s. This bit stream is continuously generated by all 52 ADCs, creating an overall data rate of 22.5 M bits/s transferred to the NPCB board through the four vertical bus connectors.

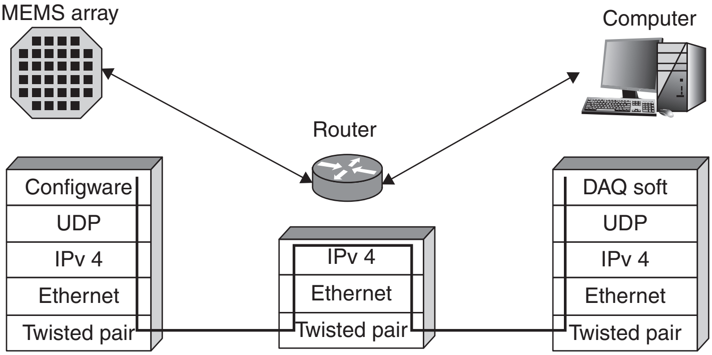
17.5 System implementation of AMA and compact and programmable data acquisition nodes (CAPTAN)

The SASA system (see Fig. 17.3) developed in this project is a novel combination of PC-/FPGA-based data acquisition system (CAPTAN) with an embedded MEMS-based microphone array (AMA). The NPCB board, as part of the CAPTAN, contains a VIRTEX-4 XC4VFX12 FPGA where all data from the array is stored and processed. The FPGA is connected to a 32

MB EPROM that contains specially designed configware for dealing with the data coming from the array. This configware is automatically loaded to the FPGA every time the system is powered up. It is important to mention that the configware plays a central part in the system architecture and it is divided in three distinct modules: (1) acquisition control module, (2) signal processing module and (3) Ethernet communication module.

The acquisition control module is the block that contains the serial peripheral interface (SPI) interface to communicate with the ADCs; this block is responsible for controlling the ADCs by programming its registers; it is also responsible for receiving and formatting the data sent by the ADCs. Data coming from the acquisition control module is then sent to the signal processing block. The processing block is application-dependent and usually contains digital signal processing (DSP) modules; however, it can be as simple as a buffer. The data is forwarded to the Ethernet block from the processing block. The Ethernet communication block is part of the CAPTAN architecture and is specially designed configware that formats data to UDP protocol (Turqueti *et al*., 2008) and sends it to the GEL board. This block does all network communication using 'User Datagram Protocol (UDP), and is a full duplex system able to transmit and receive data at a gigabit rate (see Fig. 17.6). Data are transmitted out of the FPGA through eight lines at a 125 MHz clock. These transmission lines connect to the GEL card which electrically formats the data to conform to the Ethernet physical layer using a physical layer converter (PHY). Once on the Ethernet, the information can flow directly to a computer or it can go to a network. Each GEL has a unique media access control (MAC) address and an IP (Internet Protocol) address that can be configured by the user.

An integral part of the SASA system is the data acquisition and network manager software. After the information is broadcast to the Ethernet, a



17.6 Data path and framing overview representing several layers of data encapsulation that allows the data flow through the Ethernet.

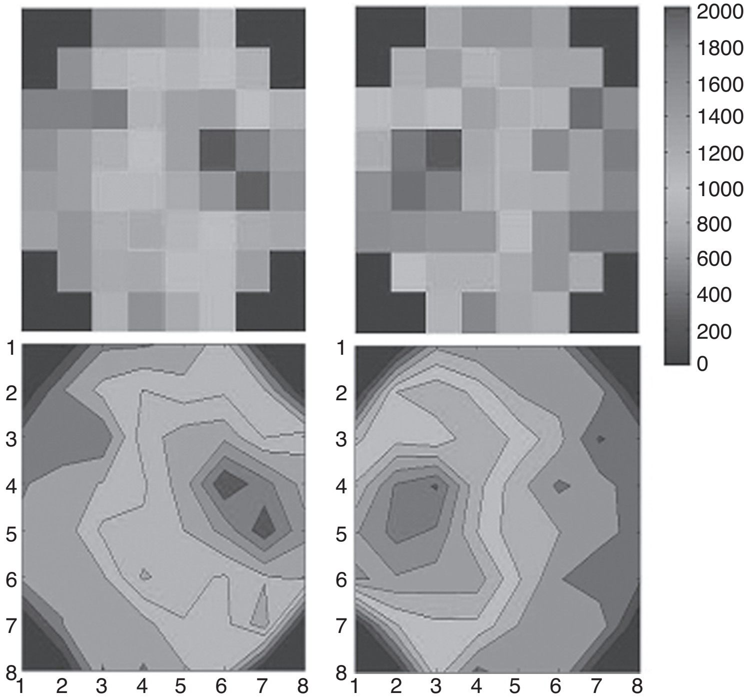
computer connected to the network is capable of retrieving the data using the CAPTAN data acquisition software. This software contains a custom class created for the SASA system, which is able to interface with the array, programming it by sending commands through the Ethernet. Commands sent by the software are interpreted by the NPCB and forwarded to the array. The software is also capable of processing and displaying the data broadcasted by the SASA in real time due to the wide link bandwidth. Due to the scalability of the SASA system, there is the option of networking multiple microphone array boards and creating a highly versatile smart acoustic sensing system.

17.6 SASA system operation

The SASA system operation is controlled by a graphical user interface (GUI) software, specifically designed for the sound sensor array based on the CAPTAN architecture. This software has the capability of programming the hardware of the microphone array with the following parameters: sampling rate, ADC sampling rate and number of sensor elements to be read. At system level, the GUI can select the board IP's with which the user wishes to establish connection. Once the GUI programs the array or arrays, the system goes into data acquisition mode. In data acquisition mode, the boards continuously send data to the computer and the computer pipes this data to a file on the hard disk. At the same time, a second thread of the software provides visualization of the raw data on the screen if the system is set for the raw data mode. If, alternatively, the system is set to processed data mode, the computer simply dumps the data in a file for subsequent analysis.

There are three modes of operation that the system can run: raw mode, processed mode and mix mode. The raw mode simply sends the following information to the computer in a 64-bit word: board ID, microphone number and 12-bit acquired data. The processed mode, on the other hand, is dependent on the configware designed by the user. It can vary from a simple signal averaging to that of a fast Fourier transform (FFT) to more complex analysis, as desired. The mix mode transmits both raw and processed data in real time. The Ethernet communication bandwidth is 800 Mbps which by far exceeds the acquired raw data bandwidth of 22 Mbps for a 36 K samples/s. The difference between the communication and raw data bandwidths allows the simultaneous transmission of raw and processed data in real time. All network communications are invisible to the user and are managed by the CAPTAN system. The user must set the IP switches on the GEL board before any network communication can take place.

A case study to demonstrate the system operating in raw mode is illustrated in Fig. 17.7. In this test, the sound source is an omnidirectional microphone

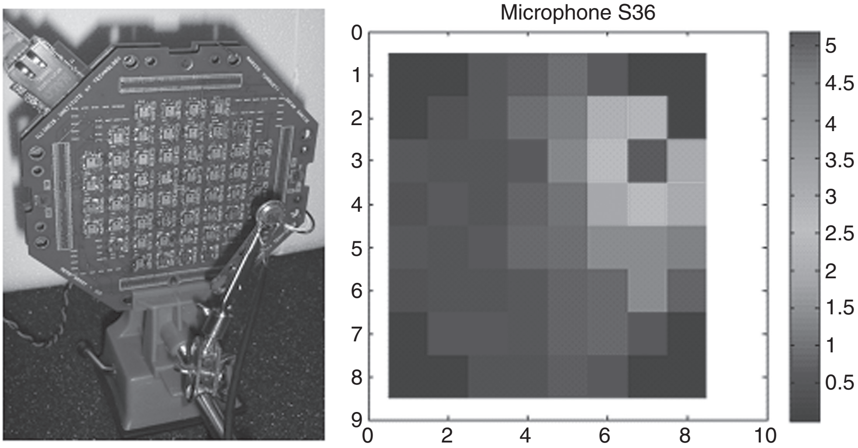


17.7 A 2D display of the system acquiring data in two different situations. The top 2D images are the raw data; the bottom 2D images are the subsequent interpolation of the raw data.

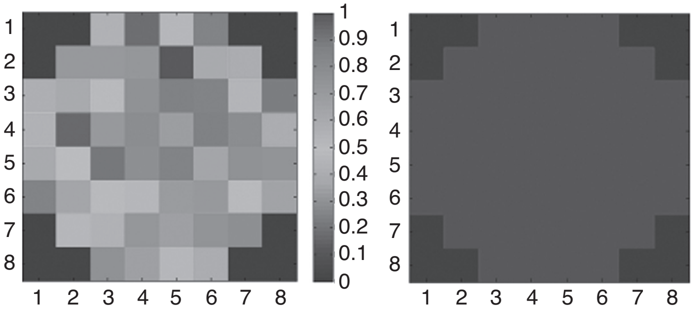
(1 cm diameter) emitting a single tone tuned at 4 kHz and located at 20 cm away from the microphone array. On the right of Fig. 17.7, the sound source is placed to the left of the center of the array board; on the left side of Fig. 17.7, the sound source is placed to the right of the center of the array board. It is also important to observe that for this test the array is not calibrated and therefore, numeric results are relative, not absolute. Nevertheless, it is possible to observe that, even with the uncalibrated raw data mode, the SASA system can obtain important information about the distribution of sound intensity.

17.7 SASA system calibration

It is necessary to calibrate the SASA system in order to conduct reliable experiments. The first step is to calibrate the individual gain of each channel on the array. The test stand used for this end is shown in Fig. 17.8 and is composed of the array itself, a CDMG13008L sound wave generator (Cui, 2006) and the clamp holder. The scale in this figure is in μPa.

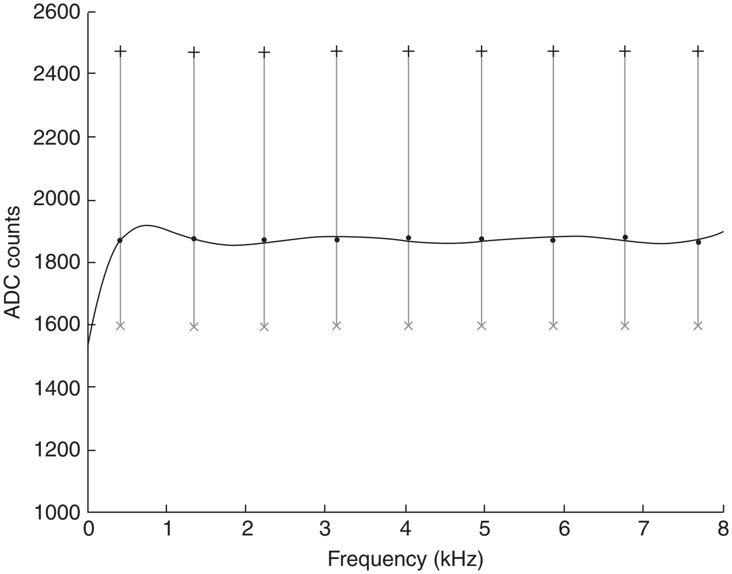


17.8 On the left the test setup; on the right is the mapping of the array responding to the stimulus of the microphone.



17.9 The gain map of the array is shown on the left; on the right is the array digitally equalized.

The next step consists of acquiring data with the array, positioning the sound source at 1 cm from the target microphone and acquiring 1 s of data. This process is repeated 52 times for the entire sensor array. The result of this process is presented in Fig. 17.9 and provides the mask for which all the data will be modified, so that the array responds homogeneously when excited. This also provides an absolute calibration of the array since at 1 cm the sound pressure was set to 5 μPa. In this case, as shown in Fig. 17.9, we have normalized the scale set from 1 to 0 where 1 corresponds to 5 μPa. The next calibration is based on a chirp waveform ranging from 100 Hz to 8 kHz to extract the frequency response of each sensor. As expected, the frequency response is homogeneous throughout the array. Figure 17.10 shows the frequency mean, minimum and maximum responses of all 52 acoustic sensors.

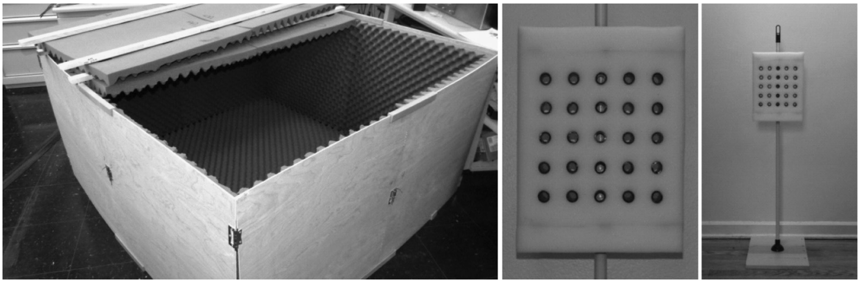


17.10 Frequency response of the array. The mean is represented by the solid line; the maximum and minimum gain are represented by the bars.

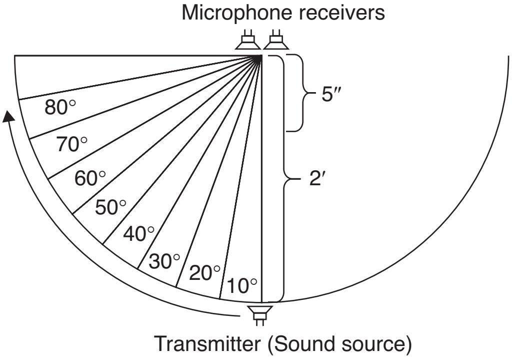
17.8 Sensor array for time-of-flight measurements

To create a controlled environment for acoustic experimentation and time-of-flight (TOF) measurements, a 52″ × 52″ × 27″ anechoic chamber was designed and built. The key features of the chamber are its ability to isolate the experiment inside the chamber from outside noise and to absorb sound inside the chamber to prevent multiple reflections (reverberation). It should be noted that while the noise from outside of the chamber could be either reflected back outside or absorbed by the chamber, sound inside the chamber has to be absorbed by the surfaces of the chamber to prevent reflections. The material used for sound absorption was a high-density 2″ thick polyester-based polyurethane convoluted foam which is specifically designed for sound absorption. Furthermore, a sensor array test stand was designed and built in order to conduct a wide variety of acoustic and ultrasound experiments. Figure 17.11 shows the outside view of the assembled anechoic chamber and a picture of the sensor array test stand.

A TOF estimation experiment was performed to determine the direction and distance of the sound source with respect to the microphones. The TOF was calculated based on the recorded phase delay and the distance between the receiving microphones. The first set of experiments involved the CAPTAN-based microphone array data acquisition system



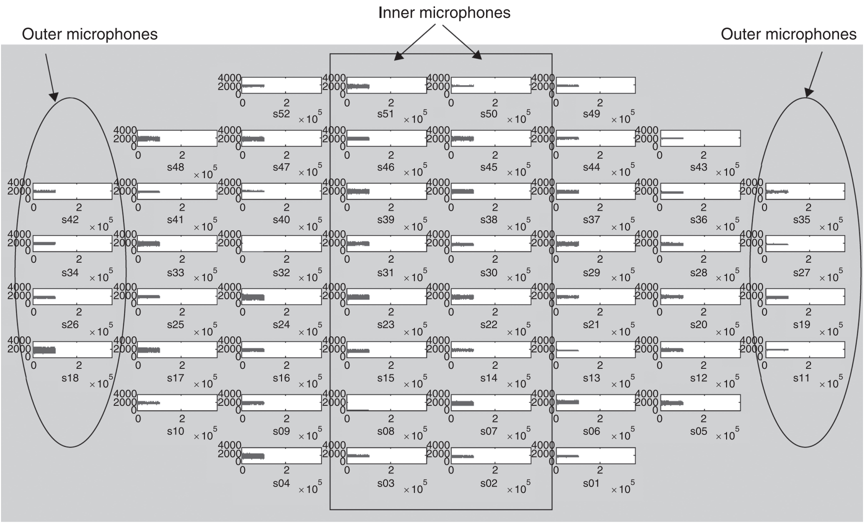
17.11 Anechoic chamber and sensor array test stand.



17.12 Sound source TOF estimation setup.

and a transmitting sound source. The microphone array data acquisition system was mounted on the vise base, the transmitting sound source was also mounted on a vise base. The experiment was carried out in a laboratory room with various random objects around the area of experimentation to create a high-noise and a highly reflective environment. The parameters varied in this experiment were the distance to the sound source, frequency of the sound source, the pairs of microphones used and the angle between the sound source and the microphones. The overall geometry of the experiment is shown in Fig. 17.12.

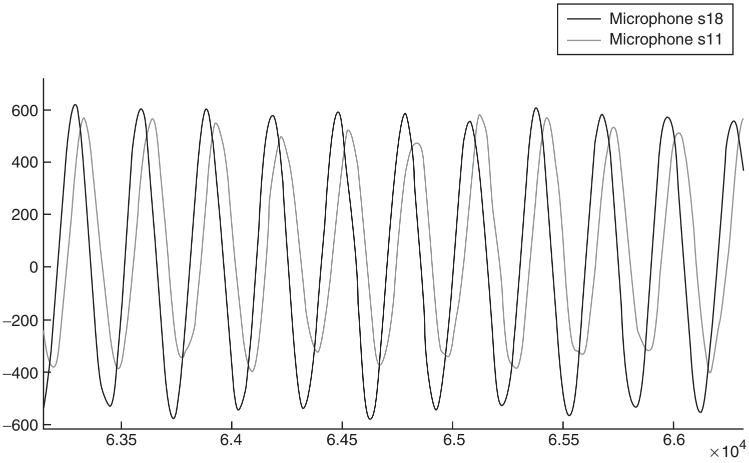
For the first test, the distance between the receivers and transmitter was 2 ft and for the second test the distance was 5 in. The sound source signal used was a continuous sine wave of frequencies 1 and 2 kHz, generated by an arbitrary waveform generator. The distance between the inner set of receivers, as shown in Fig. 17.13, was 10.0 mm, and the distance between the outer set of receivers, was 70.0 mm. The upper frequency was limited to 2 kHz to allow all the microphones of the microphone array system to be used without aliasing. The maximum source signal frequency which could



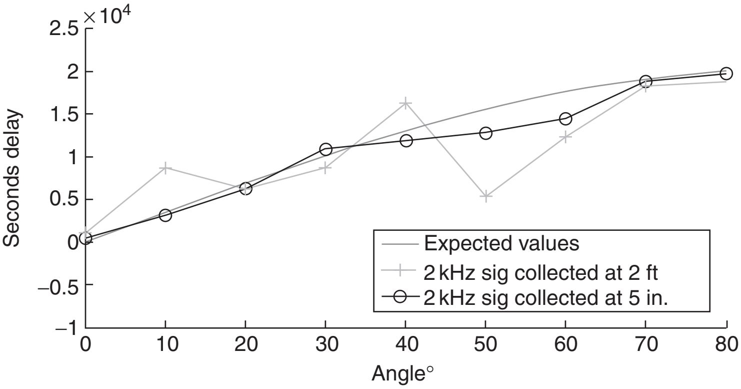
17.13 Sample data collected (horizontal lines in each microphone graph) by MEMS microphone array.

be used can be obtained from the spatial sampling theorem. This theorem states that, for a given maximum temporal frequency in the source signal, there is a minimum spatial sampling – that is there is a maximum distance between the receivers used in the acquisition system. Specifically, this maximum distance is given by $\delta \leq c/(2 * f_{\max}) = \lambda_{\min}/2$, where δ is the maximum distance between the receivers, c is the speed of sound, $f_{\max}$ is the maximum frequency of the source signal and $\lambda_{\min}$ is the minimum wavelength of the source signal.

A sample of the overall data collected by the MEMs array is shown in Fig. 17.13. In this figure, each window represents the data collected by one of the MEM microphones. The sine wave pattern of the data collected by two of the microphones is compared in Fig. 17.14. Here, the phase difference between the data collected by the two MEM microphones can be clearly seen. The sets of microphones used for experimentation were the outer set and the central inner set (see Fig. 17.13). For each set, the delay was obtained from each pair of opposing microphones; the results from each pair were then averaged to obtain the delay measurement for the set. Figure 17.15 shows a results graph comparing the expected results with those collected at 2 ft and at 5 in. for the 2 kHz source signal. It can be seen from these results that phase-based TOF in a highly noisy and reverberant environment does not produce reliable results at greater distances. This can be seen from the mismatch between the collected data and the expected results, and also from the fact that the



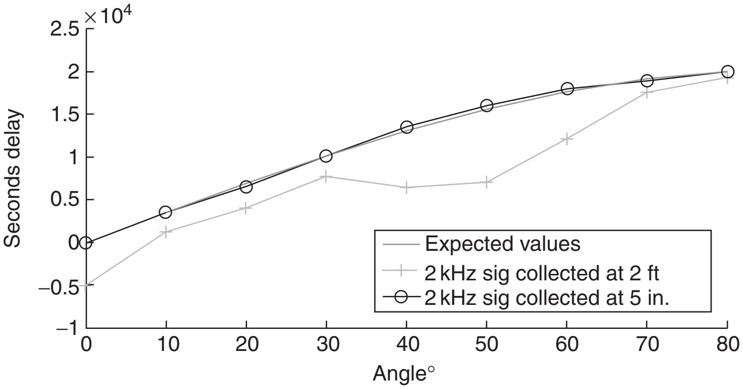
17.14 Data collected by two MEMS microphones.



17.15 TOF estimation experiment in a room environment for 2 kHz sound source at 2″ and at 5″.

time delays increase and decrease, instead of just increasing, when the angle between the sound source and receiving microphones increases. The phase-based TOF estimation works better at close distances where the power of the original signal is high compared with the power of the reflection-based noise.

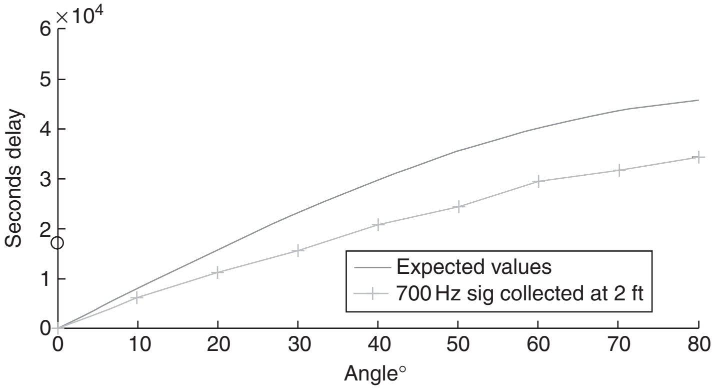
In order to reduce the effect of reflection and ambient noise, a second set of experiments was performed inside the anechoic chamber. These experiments also used the CAPTAN-based microphone array data acquisition system and a transmitting speaker. The experimental setup was the same as in the first set of experiments except for the use of the anechoic chamber.



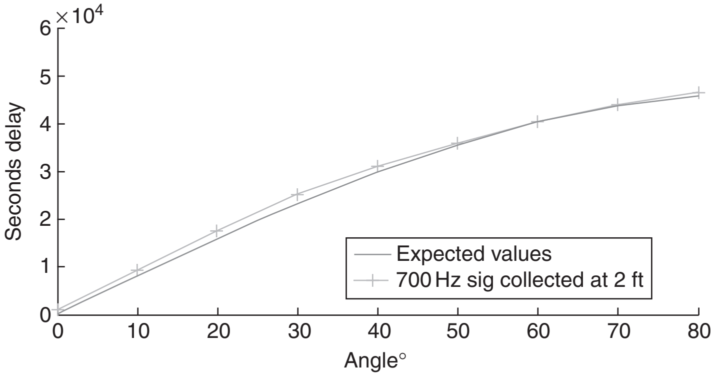
17.16 TOF estimation experiment in an anechoic chamber for 2 kHz sound source at 2″ and at 5″.

Figure 17.16 shows a results graph comparing the expected results with those collected at 2 ft and at 5 in. for the 2 kHz source signal. From these results, it can be seen that performing phase-based sound source TOF inside an acoustic chamber, which absorbs sound and thus reduces reflections, produces some improvement over performing sound source TOF estimation in a general room environment. Furthermore, the frequency of the source signal and the distance between the receiving microphones have an effect on the accuracy of the results.

Since the surrounding surfaces inside the anechoic chamber absorb most of the sound, the reflections causing the distortions in this set of experiments came from the microphone array itself. In order to reduce the effect of the reflection, an alternative physical setup and acquisition system was used for a third set of experiments. The physical setup consisted of a vise with a generic 60° beam angle microphone attached to each of the two arms of the vise by foam with nothing in between the vise arms. The distance between the microphones was increased to 6.3″. To comply with the spatial sampling theorem, the frequency of the source signals was changed to 700 Hz with the greater distance between the receivers. Figure 17.17 shows a graph comparing the expected results with those collected at 2 ft for the 700 Hz source signal. It can be observed that by increasing the distance between the microphones and removing any reflective surfaces from in between the microphones significantly improves the accuracy of phase-based TOF measurements, even at greater distances. For the 700 Hz sound signal, the measured delays follow the correct pattern but the delays are a reduced version of the expected values. This bias in estimation is again attributed to the reflection within the environment.



17.17 TOF estimation experiment in an anechoic chamber. The sensor array test stand used a 700 Hz sound source 2 ft away from the microphone.



17.18 TOF estimation experiment in an anechoic chamber. The sensor array test stand used a 20 cycle 700 Hz sine wave sound source at 2 ft away from the microphone.

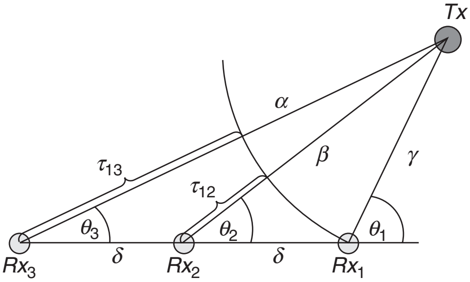
To further reduce the effects of reflection on the TOF, another set of experiments was performed. Here, the physical setup and the test parameters were the same as for the third set of experiments except for the type of sound source signal. This time, instead of using a continuous sine wave for the source signal only a 20 cycle sine wave was transmitted. Then, only the first cycle of the received pulsed sine wave is used for the phase measurement. This is expected to further reduce errors due to reflections since the first sine cycle arrives before undesirable wave reflections, and thus the phase information should be immune to distortion. Figure 17.18 shows a graph comparing the expected results with those collected at 2 ft for the 700 Hz source signal. From these results, it can be observed that phase-based

TOF measurement which uses the phase information from the first wave pulse only provides accurate results, even at greater distances and independent of the frequency of the source signal.

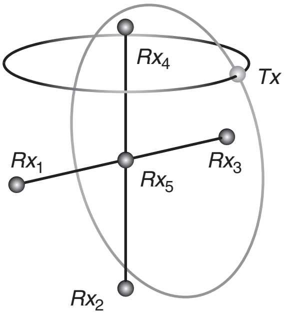
17.9 3D sound source localization

2D localization of the sound source can be performed with three receivers (Benesty *et al.*, 2008) using only the time difference of arrival (TDOA) information. The geometry for 2D localization using three receivers arranged in a line is shown in Fig. 17.19. As shown in this figure, the location of the transmitter in a plane can be obtained by measuring the TDOA between receivers 1 and 2, and also receivers 1 and 3.

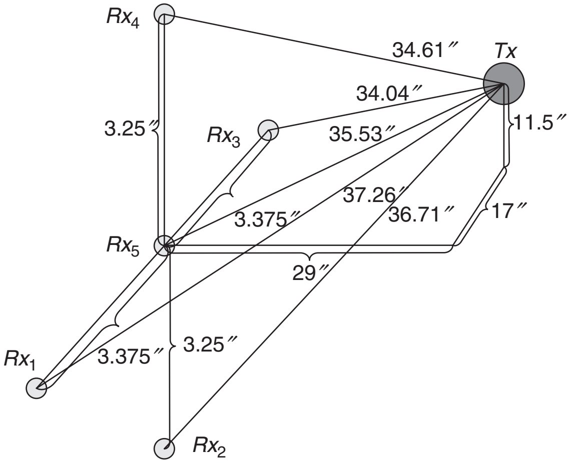
3D localization can be decomposed into two problems with 2D localizations, as shown in Fig. 17.20. Here, the transmitter is labeled *Tx*, and the receivers are labeled Rx_1 through Rx_5. Three receivers in each plane are used in the same way that they were used for 2D localization, with both results using the same coordinate system centered at receiver 5. Geometrically, this is the same as finding two circles or semicircles, the intersection of which is the location of the sound source.



17.19 2D localization geometry.



17.20 3D localization geometry.



17.21 3D distances between the receivers and the transmitter.

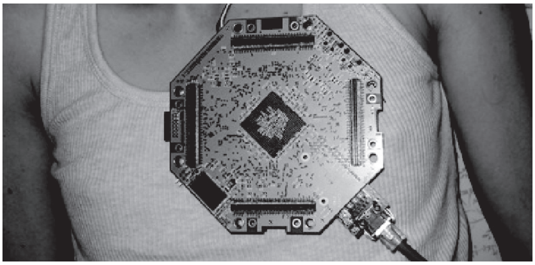
Table 17.1 Experimental results for measured 3D distances

From receiver	Collected (in.)	Expected (in.)	Error (in.)
1	37.921	37.26	0.661
2	37.378	36.71	0.668
3	34.807	34.04	0.767
4	35.213	34.61	0.603
5	36.567	35.53	1.037

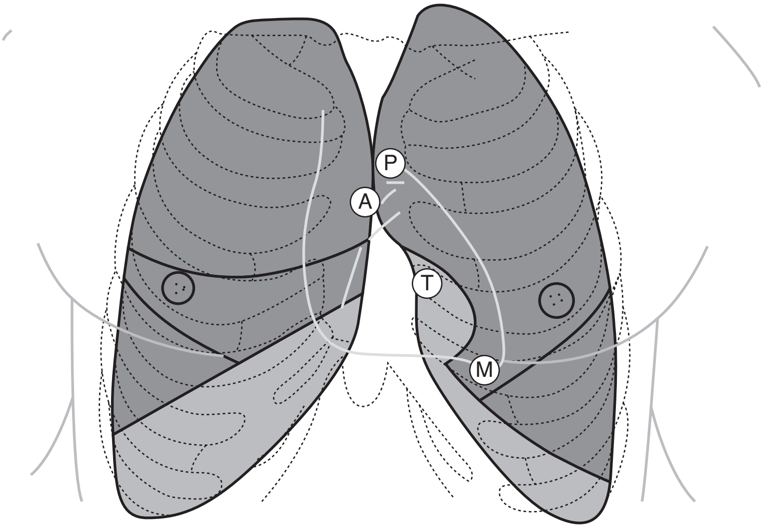
An experiment consisting of six 40 kHz ultrasonic sensors, five of which acted as receivers and one of which acted as a transmitter, was performed for 3D source localization. The transmitter used a 20 cycle sine wave pulse train. The sensor array test stand was used to hold the transmitter and receivers. The geometry and dimensions of the setup are shown in Fig. 17.21. The receivers were arranged in a plus sign pattern. The results of this set of experiments are shown in Table 17.1. As presented in the table, this experiment for 3D sound source localization produces accurate results and the collected values are within a few percent of each other.

17.10 SASA system for mapping of the heart sound

The SASA system for source separation was tested to monitor and localize the heart sound. Figure 17.22 shows the system collecting data on the chest of the subject. The system was configured with eight microphones,



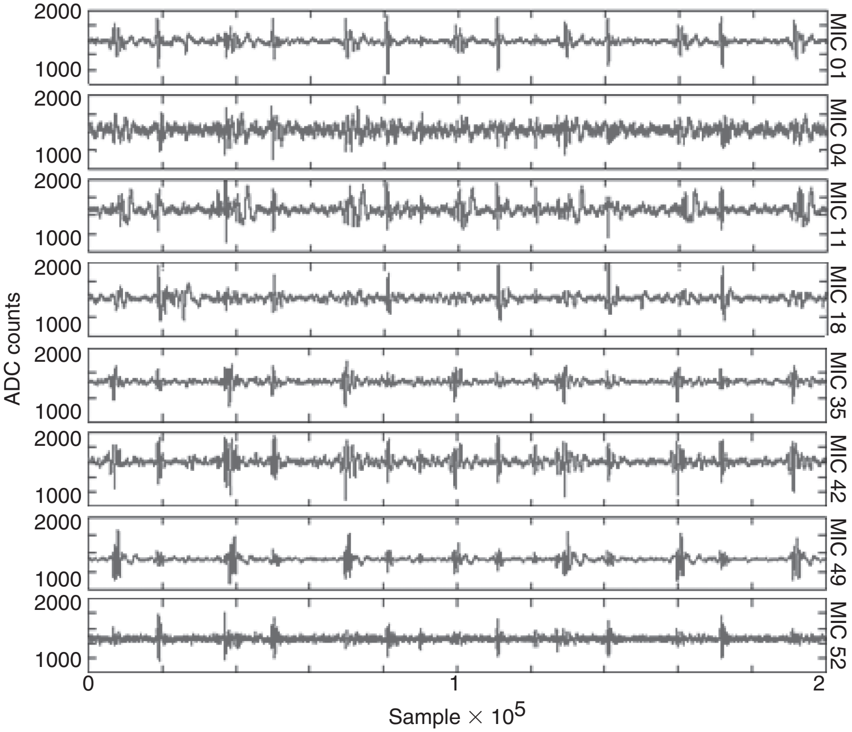
17.22 System collecting data on the subject's heart.



17.23 Localization of heart valves. See text for explanation.

with the sampling rate of 36 K samples/s for a total digitizing time of 11 s. Each heart beat in a healthy adult human is composed, basically, of two distinct sounds. The first heart sound (S1) is caused by the mitral (M) and tricuspid (T) atrioventricular valves and the second heart sound (S2) is caused by the aortic (A) and pulmonary (P) semilunar valves. The localization of these heart valves in the human chest is illustrated in Fig. 17.23 (Bates, 2005).

The results from the digitization are shown in Fig. 17.24; although 400 K samples were captured, only 200 K are displayed in this image for clarity. It is important to observe that, this time, the signals on the microphones are very different from each other. Since the array is very close to the distributed sound source due to the position of the microphones (see Fig. 17.25), they will have very different localized signals. Figure 17.26 displays sound

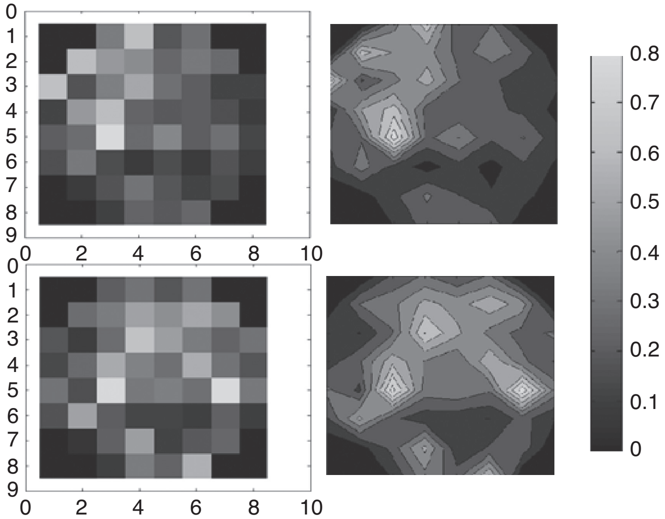


17.24 From the top to the bottom data acquired by microphones 1, 4, 11, 18, 35, 42, 49 and 52.

		52	51	50	49		
	48	47	46	45	44	43	
42	41	40	39	38	37	36	35
34	33	32	31	30	29	28	27
26	25	24	23	22	21	20	19
18	17	16	15	14	13	12	11
	10	9	8	7	6	5	
		4	3	2	1		

17.25 Numbering convention for the microphone array.

localization information collected during the heart experiment. The heart is imaged here using the sound intensity measured at each microphone; the data were subsequently interpolated and are illustrated in the same image. It is possible to observe two very distinct patterns for S1 and S2.



17.26 Sound imaging of the heart: On the top, the sound image captured when the heart beat was at the S1 stage; on the bottom, when the heart beat was at the S2 stage. The 2D images on the right are the subsequent interpolation of the raw 2D images on the left.

17.11 Conclusion

This work describes the design, development and capabilities of a scalable SASA system that can be used to enhance sound processing and acquisition for a diverse set of applications. The system uses MEMS microphone arrays in combination with the CAPTAN scalable architecture in order to deliver a powerful real-time signal processing and data acquisition platform. This research demonstrates that integration of key technologies such as MEMS sensors, high-performance FPGA devices and Gigabit Ethernet can produce a very compact, network-enabled, versatile acoustic array with a very high level of performance applicable to a broad range of applications. The SASA platform can benefit many areas of acoustic signal processing – specifically, multiple source separation, mapping and localization. Other possible applications include digital cardiac auscultation, structural health-monitoring using acoustic emission, robotic auditory systems, voice-based man–machine interfacing and ultrasound beacon-based object tracking.

17.12 References

Alghassi, H. (2008) 'Eye Array Sound Source Localization', Doctoral dissertation, University of British Columbia, Vancouver.

Bates, B. (2005) *A Guide to Physical Examination and History Taking*, 9th edn, Philadelphia, PA Lippincott Williams & Wilkins.

Benesty, J., Chen, J. and Huang, Y. (2008) *Microphone Array Signal Processing*, 1st edn, Berlin, Springer, Vol. 1.

Brandstein, M. and Ward, D. (eds) (2001) *Microphone Arrays Signal Processing Techniques and Applications*, Berlin/Heidelberg, Spring-Verlag.

Buck, M., Haulick, T. and Pfleiderer, H. (2006) 'Self-calibrating microphone arrays for speech signal acquisition: a systematic approach', *Applied Speech and Audio Processing*, **86**(6), pp. 1230–1238.

Campbell, D.K. (1999) 'Adaptive Beamforming Using a Microphone Array for Hands-Free Telephony', Master's Thesis, Bradley Department of Electrical and Computer Engineering.

Chen, B. (2002) 'Source localization and beamforming', *IEEE Signal Processing Magazine*, **19**(2), 30–39.

CUI Inc. (2011) 'CDMG13008L-02 Micro Dynamic Speaker Datasheet' http://www.cui.com/product/components/speakers/low-profile:--10~29-mm/cdmg13008l-02.

Eckert, J., German, R. and Dressler, F. (2011) 'An indoor localization framework for four-rotor flying robots using low-power sensor nodes', *IEEE Transactions on Instrumentation and Measurement*, **60**(2), pp. 336–344.

Hodgkiss, W. (1980) 'Dynamic Beamforming of a Random Acoustic Array', IEEE Acoustics, Speech and Signal Processing, *IEEE International Conference on ICASSP '80*, April 1980, Denver Colorado, Vol. 5, pp. 311–314.

Kim, S.J. and Kim, B.K. (2011) 'Accurate hybrid global self-localization algorithm for indoor mobile robots with two-dimensional isotropic ultrasonic receivers', *IEEE Transactions on Instrumentation and Measurement*, **60**(10), pp. 3391–3404.

Knowles Acoustics (2006) 'SPM204 and SPM208 Thin Amplified Mini SiSonic Microphone Specification Datasheet Rev C'.

Kunin, V., Cardoso, B., Saniie, J. and Oruklu, E. (2010) 'Acoustic Sensor Array for Sonic Imaging in Air', *Proceedings of the IEEE Ultrasonics Symposium*, 11–14 October 2010, San Diego, CA, pp. 1833–1836.

Kunin, V., Jia, W., Turqueti, M., Saniie, J. and Oruklu, E. (2011) '3D Direction of Arrival Estimation and Localization using Ultrasonic Sensors in an Anechoic Chamber', *Proceedings of the IEEE Ultrasonics Symposium*, 18–21 October, Orlando, Florida, pp. 756–759.

Llata, J.R., Sarabia, E.G. and Oria, J.P. (2002) 'Three-dimensional robotic vision using ultrasonic sensors', *Journal of Intelligent and Robotic Systems*, **33**(3), pp. 267–284, Berlin/Heidelberg, Spring-Verlag.

Nakadai, K., Okuno, H.G. and Kitano H. (2002) 'Real-Time Sound Source Localization and Separation for Robot Audition', *IEEE International Conference on Spoken Language Processing*, September 2002, Denver, Colorado, pp. 193–196.

Nishitani, A., Nishida, Y. and Mizoguch, H. (2005) 'Omnidirectional Ultrasonic Location Sensor', *IEEE Sensors Conference*, October 2005–November 2005, Irvine, CA. (pp. 4).

Rabinkin, V.D., Renomeron, J.R., French, C.J. and Flanagan, L.J. (1996) 'Estimation of Wavefront Arrival Delay Using the Cross-Power Spectrum Phase Technique', *132nd Meeting of the Acoustic Society of America*, 4 December 1996, Honolulu, Hawaii.

Rivera, R., Turqueti, M. and Prosser, A. (2008) 'A Software Solution for the Control, Acquisition, and Storage of CAPTAN Network Topologies', *IEEE Nuclear Science Symposium Conference Record*, 19–25 October 2008, Dresden, Germany, pp. 805–808.

Texas Instruments (2010) 'ADC121S101/ADC121S101–Q1 Single Channel, 0.5 to 1 Msps, 12-Bit A/D Converter', Datasheet', http://www.ti.com/product/adc121s101.

Turqueti, M., Rivera, R., Prosser, A., Andresen, J. and Chramowicz, J. (2008) 'CAPTAN: A Hardware Architecture for Integrated Data Acquisition, Control and Analysis for Detector Development', *IEEE Nuclear Science Symposium Conference Record*, 19–25 October 2008, Dresden, Germany, pp. 3546–3552.

Weinstein, E., Steele, K., Agarwal, A. and Glass, J. (2004) 'A 1020-Node Modular Microphone Array and Beamformer for Intelligent Computing Spaces', MIT/LCS Technical Memo, MIT-LCS-TM-642.

Index

AC interferometer, 235
acceleration noise power density, 374
acceleration sensor, 298
acoustic MEMS array (AMA)
  design and implementation, 496–500
    data path and framing representing several layers of data encapsulation, 501
    front-end electronics of AMA array and two-stage amplifier circuit, 499
    functional block diagram of smart acoustic sensor array (SASA) system, 499
    sensitivity comparison of two acoustic MEMS microphones, 500
    top view of CAPTAN readout system with AMA board hosting 52 MEMS microphones, 497
active pixel sensor (APS), 462–3
AD7745, 28
adaptable sensor device, 20
adaptive sensor device, 18
additive techniques, 288
ADS1232, 28
Advanced eXtensible Interface (AXI), 213
advanced microcontroller Bus Architecture (AMBA), 213
ADXL103, 28
ADXL312, 28
aluminium nitride (AlN), 107
amorphous, 284
amplitude modulation (AM), 86
amplitude ratio, 354
analog-to-digital converter (ADC), 4, 498
analogue sensors, 29
analogue sun sensor, 461–2
  angle as function of sensor's response R1 and R2, 462
  basic structure, 461
anechoic chamber, 509
anisotropic crystalline etch simulators (ACES), 293
anisotropic magneto-resistance (AMR), 390
  readout electronics effects, 394–5
  thermo-mechanical noise and intrinsic resolution, 393–4
  working principle and mechanical sensitivity, 390–3
    z-axis MEMS magnetometer, 391
anodic bonding, 295–6
ANSYS, 297, 363
application programming interface (API), 209
application-specific integrated circuit (ASIC), 59
application-specific standard processor (ASSP), 59
artificial olfactory systems (AOS), 200
atomic force microscopy (AFM), 331

beam steering, 479
beamforming, 495–6
bent beam actuator, 329
BESSY II, 112
bidirectional vertical actuator (BDVA), 326

bimorph, 317
bimorph actuator, 331
  SEM images of initial fabricated cascaded bimorph mirror, plan and side view with residual stress, 332
  Ti-W and SiN are shown to be capable of sustaining large deflection without delamination, 332
birefringent materials, 132
Block RAMS (BRAM), 221
Boltzmann's constant, 37–4
bonding techniques, 294–7, 295
  anodic bonding, 295–6
  eutectic bonding, 296–7
  fusion bonding, 295
boundary conditions
  initial conditions, 300–1
    examples of BCs and InCs, 300
Brewster angle, 149
Brownian noise, 373

Cachelink PIM, 213
capacitive displacement sensor interface
  capacitance-to-digital converter with offset capacitance cancellation and calibration functions, 67–74
    achievable resolution of three different decimation filters, 70
    chosen conventional feed-forward loop filter topology, 70
    die micrograph, 72
    experimental set-up, 73
    measurements on capacitive sensor that is mechanically actuated at 100 Hz, 73
    simplified circuit diagram of the interface, 71
  challenges for sub-nanometer displacement measurement, 63–5
    parallel plate capacitive sensor, 64
    servo loop consisting of displacement sensor and an actuator, 64
  displacement measurement in sub-nanometer range, 63–74
  offset capacitance cancellation technique, 65–7
    capacitive sensor interface based on charge-balancing principles, 67
capacitive micromachined ultrasound transducers (cMUT), 397
capacitive sensors, 32–4
  bridge-type capacitive sensor, 34
  capacitive sensor with single topology, single topology including loss term, differential topology and bridge topology, 33
  differential capacitive sensor, 34
  lossy capacitive sensor, 33–4
  single capacitive sensor, 33
capture module, 39
carbon nanotubes (CNTs), 483
central processing unit (CPU), 35
charge-coupled device (CCD), 125, 462–3
chemical vapour deposition (CVD), 287, 290–1
chevron, 317
chevron actuator, 329
circuit architectures, 478–9
circular polarised wave, 130
CLOCK_MONOTONIC, 226
clutter, 216
coarse sun sensors, 463
coefficient of thermal expansion (CTE), 295, 331
comb-drive actuators, 317–23
  2-axis MEMS scanner and electrostatic microactuator with spade-like comb structures, 324
  comb actuators illustrations, 318
  displacement enhancement, 321
  increment of displacements, 319
  laterally driven comb-drive with different suspension systems, 320
  nanopositioning stages, 322
  solid model of nanomanipulator with amplification mechanism driven by comb actuators, 323
comb-drive structure, 310–11
comb-finer actuators, 381
comb-finger actuation, 384

compact and programmable data acquisition nodes (CAPTAN), 500–2
complementary metal-oxide-semiconductor (CMOS), 314, 494
  imaging system, 125
compliant mechanism, 334–6
  SEM images of 3DOF planar micro-manipulator and topology-optimised microgripper, 336
computer aided design (CAD), 283
COMSOL, 297
continuous cone-jet mode, 411
convergence analysis, 299
coordinate measuring machine (CMM), 244
Coriolis frame, 380–1, 383
CoventorWare, 297
crab-leg flexure, 320
crystalline, 284
current-voltage conversion, 466

damping ratio, 361
data acquisition, 501–2
  system, 495
data-redundant sensor device, 20
datchik, 5–6, 17
DC interferometer, 233
Deal-Grove model, 291
deep reactive ion etching (DRIE), 293–4
degree-of-freedom (DOF), 321
  model, 351–2
degree of linear polarisation (DOLP), 135
depolarisation, 133
deposition, 290–2
  chemical-deposition process, 290–1
  physical-deposition process, 291–2
diamond, 107
differential capacitative readout technique, 392
differential capacitative sensing, 372–3
diffused reflection, 126–7
digital sensors, 29
digital signal processing (DSP), 205, 501
digital sun sensor, 462–3
  usual accuracy and field of view (FOV) of digital and analogue sun sensors, 463
digital-to-analogue converter (DAC), 69
diode region, 434
direct bonding *see* fusion bonding
direct digital interface circuits, 27–59
  applications, 53–8
    commercial microcontrollers used to implement the direct interface circuits, 53
    commercial sensors measured by direct interface circuits, 53
    experimental results of the circuit when measuring a humidity sensor, 57
    experimental results of the circuit when measuring a magnetic sensor, 56
    experimental results of the circuit when measuring a temperature sensor, 54
    experimental results of the circuit when measuring an accelerometer, 58
    experimental results of the circuit when measuring potentiometric sensor, 55
    magnetic field measurement, 55–6
    other applications, 57–8
    position measurement, 54–5
    relative humidity measurement, 56–7
    temperature measurement, 53–4
    tilt measurement, 57
  classic block diagram of sensor electronic interface, 28
  future trends, 58–9
  interface circuits, 40–53
    basic direct interface circuit for a resistive sensor, 42
    basic direct interface circuit for capacitive sensor, 43
    circuits for capacitive sensors, 47–53
    circuits for resistive sensors, 43–7

        direct interface circuit for a bridge-type resistive sensor, 47
        direct interface circuit for a differential resistive sensor, 46
        direct interface circuit for a single capacitive sensor, 49
        direct interface circuit for a single resistive sensor, 45
        direct interface circuit for bridge-type capacitive sensor, 52
        direct interface circuit for differential capacitive sensor, 51
        direct interface circuit for lossy capacitive sensor, 50
        operating principle, 40–3
        pins configuration and discharge time for each of the measurements carried out by the circuit, 45, 46, 48, 49, 50, 51, 52
        RC circuit, charging process and discharge process of RC circuit, 40
    microcontrollers, 34–40
        common specifications of oscillators used as reference in microcontrollers, 39
        general description, 35–6
        time-interval measurement, 36–40
        time-interval measurement carried out by a μC, 37
    sensors, 29–34
        capacitive sensor with single topology, single topology including loss term, differential topology and bridge topology, 33
        capacitive sensors, 33–4
        output signal of analogue sensor, quasi-digital sensor and digital sensor, 30
        resistive sensor with single topology, differential topology and bridge topology, 31
        resistive sensors, 30–2
direct electrohydrodynamic print head, 411
direct interface circuit, 28
direct methanol fuel cell (DMFC), 456–7
displacement interferometers, 231–42
    basics, 231–7
        comparison of relevant signals of a displacement, 237
        heterodyne interferometers, 234–6
        homodyne interferometers, 233–4
        principle of heterodyne laser interferometer, 234
        principle of homodyne laser interferometer, 233
        signals, 236–7
    interferometer concepts, 237–9
        linear interferometer, 237
        optics in linear interferometer, 238
        optics in plane mirror interferometer, 238
        plane mirror interferometer, 237–9
    phase detection and interpolation, 239–42
        2D interpolation using 2D lookup with pre-calculated arctangent values, 241
        phase interpolation using direct arctangent calculation, 241
        signal generation by thresholding and phase interpolation, 242
    schematic representation of Michelson's interferometer, 231
displacement sensors
    MEMS integrating motion, 366–97
        future trends, 395–7
        MEMS gyroscope, 380–90
        MEMS magnetometer, 390–5
        technical description of MEMS motion sensors, 369–80
Doppler shift, 236
double refraction, 132
dripping mode, 410
driving MEMS switches, 447–9
    PSpice simulation results of photovoltaic source and MEMS switch models, 449
drop-on-demand (DOD), 413
droplet ejection modes, 409–11

dry etching, 293–4
dynamic analysis, 353
dynamic element matching (DEM), 69
dynamical behaviour
  smart MEMS in industrial applications, 349–64
    quality factor and loss of coefficient of vibrating structures, 358–63
    resonant accelerometers example, 364
    resonant frequency response of vibrating structures, 350–8

eddy-current sensors
  design requirements in precision industrial applications, 84–6
    ECS system with functional blocks, 85
  harsh industrial environments, 76–99
    simplified differential configuration of ECS sensors for measuring gas pressure, 77
  interfaces with an LC oscillator and ratiometric measurement, 89–99
    conceptual illustration of the precision peak detection based ECS system, 92
    interface using bandpass ΔΣ converter, 95
    LC front-end oscillator with a single resonator, 90
    precision peak-detection-based ECS interface, 90–4
    precision peak detection principle, 91
    qualitative comparison for alternative ECS interface design methods, 98
    simplified circuit diagram of synchronous detection-based ECS interface, 97
    synchronous-detection based ECS interface, 96–9
    synchronous detection-based interface with a system level compensation for the offset originated from self-mixing effect, 95
    trade-offs in mixer-based interfaces, 94–6
    transfer characteristic example with ratiometric readout function, 91
  limitations, 81–4
    frequency dependence, 84
    limited sensing-coil quality factor, 83–4
    parasitic effects, 83
    penetration depth for different materials and excitation frequencies, 81
    replacing the target with penetrating eddy current by a lossless image plane, 81
    skin effect, 81–3
  principles of operation and practical limitations, 78–84
    measured inductance and series resistance of flat sensor coil, 80
    operating principle of an ECS, 79
    sensor operation principles, 78–80
  state-of-the-art interfaces, 86–9
    functional block diagram of the interface using SC oscillator for sensor excitation, 88
    interface implemented in Kejik et al, 89
    relaxation-oscillator based interface, 88–9
    standard functional diagram of ECS interface, 86
    utilising external switched-capacitor oscillator and LC resonator, 87–8
Edlén equation, 268
  updated, 249
effective number of bits (ENOB), 194
electrical behaviour, 433
electrical module, 301
electrical resistance, 154
electrical resistivity, 154
electrical stiffness, 377
electrodeposition, 291
electrodes, 450

electrohydrodynamic (EHD) print head droplet ejection, 407–13
  configuration, 411–13
    three types of EHD print head, 412
  principle, 407–9
    schematic of EHD spray phenomenon, 408
  various droplet ejection modes, 409–11
    dripping mode, micro-dipping mode, pulsed cone-jet mode and cone-jet mode, 410
    EHD ejection mode by electric potential, 409
electrohydrodynamic (EHD) printing applications
  case study, 423–7
    image and array achieved by EHD jet printing, 425
    images of patterns of DNA formed by e-jet printing, 427
    microscope images of fine patterns using silver Nanopaste, 424
    microscopic nozzle structure of tilted-outlet nozzle, 426
electrohydrodynamic (EHD) smart printing system, 413–23
  control of pattern resolution and thickness, 417–19
    patterns of hydrophobic and hydrophillic substrates and jetting image of ink, 419
    relationship between patterned dot size and nozzle size and patterning result, 418
  multi-head printing system, 422–3
    effect of meniscus by electrical interference, 422
    gate control electrode and direct EHD multi-head printing system, 423
  printing system, 413–17
    commercialised printing systems by ENJET Co., 417
    drop-on-demand ejection characteristics using high voltage supplier, 414
    high voltage supplier, 414
    jetting images of drop-on-demand results using conductive nanosilver ink, 415
  repeatability and stability, 419–22
    improvements in repeatability and stability, 421
    moment of ejection and meniscus shape images, 420
electrolyte, 450
electromechanical behaviour, 484–5
electronic readout, 374–8
electronic sensitivity, 377–8
electroplating, 291
electrospray, 407–8
electrostatic actuation, 309–10
electrostatic force, 309, 376
electrostatic systems, 309–12
  common features, 310–12
    collapse of serpentine flexure due to short circuit between electrodes, 313
    microactuator with uniform comb structure and spade-like comb structure, 311
    optical microswitch with curved electrode structure, 311
    SEM of electrostatic micromirror with gap-closing actuation, 311
  electrostatic actuation, 309–10
electrothermal compliant mechanism (ETCs), 334–6
electrothermal systems, 314–17
  common features, 315–17
    cascaded bent beam actuator, 330
    response time of electrothermal microactuator with large vertical displacement, 316
    SEM image of electrothermal MEMS scanner, 320
  electrothermal actuation, 315
embedded structure-based polymeric electrothermal actuators, 330–31
epitaxy, 291

error detector, 16
etching, 292–4
  dry, 293–4
  techniques, 288, 292
  wet, 293
    simulations of etching processes using ACES for different wafers, 293
Ethernet communications, 498
eutectic bonding, 296–7
evaporation, 292
extinction ratio (ER), 140
extreme ultraviolet (EUV)
  advanced silicon radiation detector, 102–20
    challenges for radiation detection in the EUV and VUV spectral ranges, 103–7
    device solutions for radiation detection in EUV and VUV spectral ranges, 107–11
    future trends, 120
    methods of radiometric investigation and characterisation, 111–13
    spectral responsivity and radiation hardness of EUV and VUV radiation detectors, 113–20

fabrication process, 454
fast Fourier transform (FFT), 502
fault-tolerance, 19
field assisted bonding *see* anodic bonding
field of view (FOV), 460
field programmable gate arrays (FPGAs), 205
fill factor, 436
fine sun sensors, 463
finite element method, 334–6
flexible panel displays (FPDs), 424
flip-chip technology, 441–2
fluid structure interaction (FSI), 301
folded-beam designs, 320
frequency modulation (FM), 86
Fresnel reflection theory, 144
full-scale range (FSR), 373
fusion bonding, 295
  technology, 443–5

gallenium arsenide (GaAs), 107
gate electrode electrohydrodynamic print head, 411–12
Gauss's Law, 312
general-purpose interrupt, 38–9
Gigabit Ethernet Link (GEL), 498
glare, 127
global positioning system (GPS), 457
gold diode, 104
good calibration, 466
graphical user interface (GUI) software, 502

hardware-centered, 8
heart sound mapping, 512–15
helium-neon (Hene) lasers, 230
heterodyne interferometers (detection), 234–6
Heydemann correction, 239, 268
high-isolation X-band MEMS capacitive switches, 475–6
homodyne interferometers (detection), 233–4
Hooke's Law, 312
Huygens-Fresnel principle, 257

industrial applications
  dynamical behaviour of smart MEMS, 349–64
    quality factor and loss of coefficient of vibrating structures, 358–63
    resonant accelerometers example, 364
    resonant frequency response of vibrating structures, 350–8
  microfabrication technologies for creating smart devices, 281–303
    materials, 283–7
    MEMS design and modelling, 282–3
    processes, 287–97
    simulation, 297–302
industrial printing
  microelectromechanical systems (MEMS) print heads, 402–28
    case study of EHD printing applications, 423–7

    - EHD print head droplet ejection, 407–13
    - EHD smart printing system, 413–23
- inertial measurement units (IMU), 390
  - combined with proximity sensors, 396–7
    - illustration based on changes of fringe capacitance, 397
- infrared (IR) imaging, 396
- Institute of Electrical and Electronics Engineers (IEEE), 5
- integrated circuit (IC) technology, 287
- integrating motion
  - MEMS displacement sensors, 366–97
    - future trends, 395–7
    - MEMS gyroscope, 380–90
    - MEMS magnetometer, 390–5
    - technical description of MEMS motion sensors, 369–80
- intelligent sensor, 8
  - device, 19
  - hub, 9
- interdigitated fingers, 310–11
- interface circuits, 40–53
  - circuits for capacitive sensors, 47–53
    - bridge-type capacitive sensor, 52–3
    - differential capacitive sensor, 51–2
    - lossy capacitive sensor, 49–50
    - single capacitive sensor, 48–9
  - circuits for resistive sensors, 43–7
    - bridge-type resistive sensor, 46–7
    - differential resistive sensor, 45–6
    - single resistive sensor, 44–5
  - operating principle, 40–3
- Internet Protocol (IP) address, 501
- intrinsic damage, 117
- intrinsic device noise, 387–9
- intrinsic resolution, 393–4

- Joule effect, 393–4
- Joule heating, 315

- Laplace equation, 370
- light dependent resistors (LDRs), 156
- linear polarised wave, 130
- liquid atomisation mode, 407–8
- Lissajous curve, 239
- Lorentz force, 78
  - capacitative magnetometers, 393
- low-noise amplifier (LNA), 487
- low pressure chemical vapour deposition (LPCVD), 291, 454

- machine vision
  - application of polarisation vision, 143–9
    - measurement set-up, 145
    - polarisation Fresnel ratio of 0° and 90° polarisation sensitive pixel, 148
    - Stokes degree of linear polarisation for 0° and 90° polarisation sensitive pixel, 149
    - theoretical Fresnel reflection coefficients for aluminium and plastic, 147
    - theoretical polarisation Fresnel ratio, 147
    - transmitted intensity at 0° and 90° polarisation sensitive pixel, 146
    - transmitted radiance for plastic and aluminium, 146
    - transmitted radiance of reflected light as function of transmission axis of the linear polariser, 145
  - design of polarisation sensor, 141–3
    - pixel cross-section and top view of implemented wire-grid polariser, 141
    - sensor regions with different polarising angles, 142
  - integrated polarisation analysing CMOS image sensors, 124–50
    - lighting arrangement in machine vision applications, 125
    - reflection types: specular and diffused reflections, 126
  - polarisation cameras, 135–41
    - comparison of linear polarisers, 139
    - polarisation characterisation and measurements, 139–41
    - polarisation filter types, 136–9
    - wire-grid polariser, 138
  - polarisation vision, 127–35
    - linear polarisation, 131
    - polarisation by absorption, 131

machine vision (*cont.*)
    polarisation by double refraction, 133
    polarisation by reflection, 132
    polarisation of light: basics, 128–30
    polarisation of light: representation, 133–5
    polarisation of light from an unpolarised beam, 130–3
main photovoltaic electrical parameters, 435–8
  characteristics of current-voltage I-V and power-voltage P-V, 437
Malus's law, 140
materials, 283–7
  metals, 286
  photoresists, 285–6
    negative and positive photoresists are exposed throughout a mask, 286
  piezoelectric and piezoresistive materials, 287
  polymers, 284–5
  silicon, 284
measuring system, 17
measuring transducer, 5
mechanical model, 369–72
mechanical module, 301
mechanical resonator sensors, 364
mechanical sensitivity, 373, 383–7
mechanical stiffness, 351
media access control (MAC), 501
membrane electrode assembly (MEA), 454
memory, 35
mesh model, 299–300
  accelerometer sensor, 300
metal-air-metal (MAM) capacitors, 477
metal grid micropolariser arrays, 138
metal-insulator-metal (MIM) capacitor, 476
metal oxide sensors (MOX), 156
metals, 286
metrological self-check of a sensor device, 18
metrology light source (MLS), 112
Micro and Nano Technology (MNT), 440
micro-dripping mode, 410–11
micro-fuel cells, 449–56
  applications, 456–7
  fuel-cell principles and classifications, 449–52
    basic structure, 450
    channelled structure of an electrode, 451
    properties most usually employed, 452
  MEMS-based PEM fuel cell, 452–6
    design illustration, 453
    detail of mask patterning for pillar design, 455
    fabrication process for PEM silicon fuel cell, 455
    general structure of silicon fuel cell with integrated fuel distribution chamber, 454
    measurement of power density and output voltage, 456
microactuators
  design and analysis, 317–36
    bimorph actuator, 331
    cantilever type piezoelectric actuators, 331–3
    comb-drive actuators, 317–23
    compliant mechanism, 334–6
    embedded structure-based (silicon skeleton) polymeric electrothermal actuators, 330–1
    membrane type piezoelectric actuators, 333–4
    parallel-plate actuator, 323–5
    scratch drive actuator, 324–5
    U-shaped (pseudo bimorph) beam, 326–9
    V-beam (chevron/bent-beam) actuator, 329–30
  design and technology, 305–37
    common microactuation mechanisms, 306
  driving principles for actuators, 307–18
    electrostatic systems, 309–12
    electrothermal systems, 314–17
    piezoelectric systems, 312–15
  mechanisms selection considerations, 307–9

      actuation resolution and sensing, 308
      fabrication and material selection, 308–9
      output force and displacement range, 308
      power consumption, 309
      size and physical properties, 308
Microblase processor, 207
microbridges, 350
microcantilevers, 350
microelectromechanical systems (MEMS)
   design and modelling, 282–3
      design flow of manufacturing process of microsystem, 282
   future trends, 395–7
      combination of inertial measurement units (IMUs) with proximity sensors, 396–7
   integrating motion and displacement sensors, 366–97
      MEMS gyroscope, 380–90
      MEMS magnetometer, 390–5
      specifications of MEMS accelerometers and gyroscopes for consume and automotives, 369
      technical description of MEMS motion sensors, 369–80
   microphones, 494
      pre-amplifier, acoustic transducer and aluminium cover forming resonant cavity, 494
   print heads for industrial printing, 402–28
      case study of EHD printing applications, 423–7
      EHD print head droplet ejection, 407–13
      EHD smart printing system, 413–23
microelectromechanical systems (MEMS) accelerometer, 369–80
   differential capacitative sensing, 372–3
   electronic readout, 374–8
      schematic illustration, 375
   mechanical model, 369–72
      modulus of transfer function between external force and displacement of uniaxial MEMS, 371
   resonant accelerometers, 378–80
      schematic structure of resonant accelerometer and readout electronics, 379
   schematic illustration of uniaxial MEMS accelerometer, 370
   thermo-mechanical noise, 373–4
microelectromechanical systems (MEMS) gyroscope, 380–90
   intrinsic device noise, 387–9
      example of noise power spectral density, 389
   mechanical sensitivity, 383–7
      transfer function of Coriolis frame, 388
   reference to resonant gyroscopes, 389–90
   working principle, 380–3
      dual-mass tuning fork gyroscope, 382
      single mass gyroscope for sensing of angular rate along z-axis, 381
microelectromechanical systems (MEMS) magnetometer, 390–5
microfabrication technologies
   creating smart devices for industrial applications, 281–303
      materials, 283–7
      MEMS design and modelling, 282–3
   processes, 287–97
      bonding techniques, 294–7
      deposition, 290–2
      etching, 292–4
      photolithography, 289–90
      schematic of generic manufacturing process flow, 288
   simulation, 297–302
      boundary and initial conditions, 300–1
      mesh model, 299–300
      simulation and fabrication materials, 297–8
      solid model, 298–9
   simulation results, 301–2

microfabrication technologies (*cont.*)
    representation of displacement of accelerometer microstructure, 302
    representation of Von Mises's stresses of accelerometer microstructure, 302
micropolariser array, 137
MicroWires, 138
mini-modules technologies, 439–45
  fusion-bonding technology, 443–5
    2" processed wafer with different cell topologies before deposition, 444
    mini-module of nine cells in series processed, 444
  multi chip module (MCM) technology, 440–2
    sketch of MCM mini-module concept and fabricated module before dicing, 441
  silicon-on-insulator (SOI) technology, 442–3
    top view of SOI cell and fabricated nodule of 81 solar cells, 443
monolithic integration, 486, 487
monolithic mini-module fabrication, 445
monomers, 284
motion equation, 369–70, 377
multi chip module (MCM) technology, 440–2
Multi-User-MEMS-Process (MUMPs), 317
multiparameter optimisation algorithm, 334–6
Multiphysics method, 363

nafion, 451
nanoelectromechanical systems (NEMS), 322
National Institute of Standards and Technology (NIST), 112
National Metrology Institute of Japan (NMIJ), 112
National Metrology Institutes (NMIs), 112
Navier-Stokes equation, 283
negative temperature coefficient (NTC), 156
network manager software, 501–2
node processing and control board (NPCB), 498
noise transfer function (NTF), 96
nonlinearities, 247
Nyquist-Shannon sampling theorem, 496

offset capacitance, 65
Ohm's Law, 154, 161, 165
operational amplifier (opamp), 66
optic axis, 132
optical incremental sensors
  current and future trends, 273–4
  design considerations, 267–73
    Abbe correction, 269–70
    encoder-based system with scales oriented be collinear with the work point to eliminate Abbe errors, 269
    grating scale errors, 268
    measurement direction for single symmetric encoder, 272
    multi axis encoder systems, 271–3
    periodic errors, 268
    separate X and Y-encoder, 271
    stability, 267–8
    thermal expansion, 270–1
    X, Y, Rz encoder configuration with thermal centre., 270
    XZ encoder concept based on integration of two tilted encoder heads, 272
  displacement interferometers, 231–42
    basics, 231–7
    determining two quadrature signals from linear combination of sensor signals, 241
    Heydemann correction and Lissajous curve, 240
    interferometer concepts, 237–9
    interpolation using 2D lookup with pre-calculated arctangent values, 241
    optics in linear interferometer, 238
    optics in plane mirror interferometer, 238

  phase detection and interpolation, 239–42
  phase interpolation using direct arctangent calculation, 241
  principle of heterodyne laser interferometer, 234
  principle of homodyne laser interferometer, 233
  schematic representation of Michelson's interferometer, 231
  signal generation by thresholding and phase interpolation, 242
encoders and interferometers, 230–74
environment dependent errors, 248–9
  refractive index of air, 249
  thermal effects on the interferometer, 248
inferential encoders, 255–67
  diffraction physics, 257–8
  phase detection, 263–5
  practical example, 266–7
  schematic set-ups, 261–3
  sensitivities, 259–61
  tilt sensitivity, 265
instrument dependent errors, 246–8
  beam walk off, 246–7
  electronics and data age, 247
  periodic deviation, 247–8
  (split) frequency, 246
optical encoders, 249–67
  absolute linear encoder scales, 250
  basic principle of interferential grating encoder, 256
  derivation of grating equation, 258
  design requiring spatially coherent source and with low tilt tolerance, 266
  double-pass interferential grating encoder with signal period of one-quarter of the grating period, 267
  encoder with four-phase zones with 90° phase offsets, 254
  homodyne phase detector, 265
  imaging incremental encoder, 251–5
  imaging incremental encoder with source, detector, collimator lens, 252
  imaging moiré encoder, 252
  incremental encoder principles, 251
  interferential encoders using reflective scales, 264
  light diffraction of coherent light beam by amplitude grating, 257
  optical encoder with index grating and scale grating with a small pitch difference, 254
  phase sensitivity of diffracted wavefront, 259
  reflective and transmissive diffraction angles, 258
  sensitivity vectors of reflective scale grating, 260
  sensitivity vectors of transmissive scale grating, 261
  Talbot effect, 253
  transmission type encoder, reflective type encoder, phase zones, 255
  transmissive diffraction encoders and the sensitivities, 262
setup dependent errors, 243–6
  Abbe error, 243–4
  cosine error, 243
  dead path error, 244–5
  mechanical stability, 246
  target uniformity, 245
sources of error and compensation methods, 243–9
  Abbe error in an AFM measurement, 244
  cosine error occurring an interferometer measurement, 244
  dead path error and target uniformity, 245
  measurement results with periodic deviation, 248
  relevant beam parts for frequency stability effects, 247
  sensitivity of refractive index of air, 249
  ultra precision coordinate measuring machine without Abbe errors, and laser tracker, 245

optical scale, 250
order of diffraction, 257
oscillator-based systems, 175–88
  parasitic capacitance issues, 177–82
  problem of long measuring times, 182–88

parallel kinematics mechanisms (PKMs), 321
parallel-plate actuator, 323–5
  SEM images of curved electrode with stopper and insulator, slit actuator and L-shaped spring, 325
parasitic capacitance, 157–60
  effect due to sensor connections with the measurement system, 159
  effect in gas sensors with embedded heater filament, 159
  effect in sensors with interdigitated electrodes, 158
  simplified model of resistive sensor taking into account the parasitic capacitive effects, 159
Paschen's Law, 308
pass-through absolute minimisation, 214
peripherals, 35
photocathode, 103
photoconductive region, 434
photocurrent, 434–5
  definition of light incidence angle, 437
  spectral response vs wavelength for different photovoltaic semiconductor materials, 435
  typical light spectrums, 436
photoemissive detectors, 103
photolithography, 289–90
photoresists, 285–6
photovoltaic mini-generators, 433–45
  applications, 445–9
    driving MEMS switches, 447–9
    telesupplying bio-implantable devices, 445–7
  mini-modules technologies, 439–45
    comparison between mini-module fabrication technologies, 439
  working principles, 433–9
    I-V characteristics of photovoltaic cell for dark and light conditions, 433
photovoltaic modules, 438–9
  ideal module consisting of connection branches in parallel, 438
  scaling up rules of ideal module consisting of branches in parallel, 438
photovoltaic region, 434
physical layer converter (PHY), 501
physical vapour deposition (PVD), 291–2
Physikalisch-Technische Bundesanstalt (PTB), 112
piezoelectric actuation, 312–14
piezoelectric actuators
  cantilever type, 331–34
    bend-up/bend down actuator lifting load and SEM of vertical stage actuation, 333
  membrane type, 333–34
    when voltage is applied, piezoelectric actuator will push membrane slice upwards, 334
piezoelectric materials, 287
piezoelectric systems, 312–15
  common features, 313–15
    response of piezoelectric microactuator, 314
    SEM image of piezoelectric microactuator with micro-compliant amplifier, 316
  piezoelectric actuation, 312–14
    operating mode of piezoelectric beam, 313
piezoresistive materials, 287
piezoresistive module, 301
plane polarised light, 130
plasma enhanced chemical vapour deposition (PECVD), 291, 477

PLB PIM, 213
Polarcor, 136
polarisation cameras, 135–41
  comparison of linear polarisers, 139
  polarisation characterisation and measurements, 139–41
  polarisation filter types, 136–9
  wire-grid polariser, 138
polarisation ellipse, 133
Polarisation Fresnel Ratio (PFR), 146
polarisation illumination, 127
polarisation vision, 127–35
  applications, 143–9
    measurement set-up, 145
    polarisation Fresnel ratio of 0° and 90° polarisation sensitive pixel, 148
    Stokes degree of linear polarisation for 0° and 90° polarisation sensitive pixel, 149
    theoretical Fresnel reflection coefficients for aluminium and plastic, 147
    theoretical polarisation Fresnel ratio, 147
    transmitted intensity at 0° and 90° polarisation sensitive pixel, 146
    transmitted radiance for plastic and aluminium, 146
    transmitted radiance of reflected light as function of transmission axis of the linear polariser, 145
  linear polarisation, 131
  polarisation of light: basics, 128–30
  polarisation of light: representation, 133–5
  polarisation of light from an unpolarised beam, 130–3
    polarisation by absorption, 131
    polarisation by double refraction, 133
    polarisation by reflection, 131–2, 132
    polarisation by refraction, 132–3
    polarisation by scattering, 133
polarised light, 128, 130
polariser, 131
Polaroid, 131
polling, 38
polycrystalline, 284
polydimethylsiloxane (PDMS), 331
polymer electrolyte membrane (PEM) fuel cell, 452–6
polymerisation, 284
polymers, 284–5
positive temperature coefficient (PTC), 156
potentiometers, 157
power consumption, 309
power microelectromechanical systems
  photovoltaic and fuel cells for smart energy management, 431–68
    micro-fuel cells, 449–56
    micro-fuel cells applications, 456–7
    photovoltaic mini-generators, 433–45
    photovoltaic mini-generators applications, 445–9
    sun sensors, 457–68
primary standard, 111
print heads
  microelectromechanical systems (MEMS) for industrial printing, 402–28
    case study of EHD printing applications, 423–7
    EHD print head droplet ejection, 407–13
    EHD smart printing system, 413–23
    electrohydrodynamic (EHD) print head, 407
    MEMS inkjet print head vs conventional fabrication process, 403
    piezoelectric print head, 405
    thermal-bubble print head, 404
printed circuit board (PCB), 315
proximity sensors, 396–7
pseudo bimorph *see* u-shaped beam
pseudo-sine wave, 87
PSPICE model, 448
pull-in phenomenon, 377
pull-in voltage, 352, 447–8
pulsed cone-jet mode, 410–11
'pure boron' technology, 109

quasi-digital sensors, 29

Radio Corporation of America (RCA), 288
radio frequency microelectromechanical systems (RF-MEMS)
  co-integration of RF-MEMS-based circuits with integrated circuits (IC), 485–7
    placements illustration, 486
  power capability, 483–5
    measured capacitor values vs RF-power stresses intensity, 484
    microwave behaviour vs DC control voltage for different power stress levels, 485
  reliability, 482–3
    dielectric charging in capacitive RF MEMS, 483
  smart communication microsystems, 472–88
    RF-MEMS-based circuits, 478–82
    RF MEMS switches, 473
  technology and devices, 473–7
    bridge and cantilever configurations, 474
    capacitive switches, 476, 477
    capacitive varactor, 477
    technological process flow, 475
ratiometric function, 90, 92
reactive ion etching (RIE), 293–4, 454, 482
readout electronics, 394–5
real time Ethernet (RTE), 202
real-time sound processing applications
  smart acoustic sensor array (SASA) system, 492–515
    3D sound source localisation, 511–12
    design and implementation of smart acoustic MEMS array (AMA), 496–500
    fundamentals, 495–6
    heart sound mapping, 512–15
    MEMS microphones, 494
    SASA system calibration, 503–5
    SASA system operation, 502–3
    sensor array for time-of-flight measurements, 505–11
    system implementation of AMA and CAPTAN, 500–2
reconfigurable filters, 480
reconfigurable ultrasonic smart sensor platform (RUSSP)
  algorithms used in evaluation of RUSSP, 214–19
    chirplet signal decomposition, 217–19
    chirplet signal decomposition flowchart for estimation of ultrasonic echoes, 220
    coherent averaging, 215–16
    split spectrum processing, 216–17
    ultrasonic split spectrum processing algorithm, 217
  design, 209–14
    ADC to FPGA interconnection with recovered clock, 214
    analogue-to-digital converter (ADC) to FPGA interface, 213–14
    embedded system and bus interfaces showing the maximum bandwidth information, 212
    FPGA interconnection without recovered clock, 214
    RUSSP architecture, 210–13
    RUSSP block diagram showing hardware components and inteconnections, 211
    system features and user interface, 209
    system response and real-time operational requirements, 210
  fundamentals of ultrasonic sensing and pulse-echo measurements, 207–9
    ultrasound experimental setup for real-time signal processing and imaging applications, 208
    Virtex-5 FPGA, ADC, DAC, pulse/receiver and ultrasonic transducer, 208
  future trends, 226–7

  - hardware realisation of ultrasonic imaging algorithms, 219–26
    - ADC coherent averaging block, 222
    - ADC system implementation, 221
    - averaging implementation, 219–21
    - block diagram of the SSP implementation, 223
    - chirplet signal decomposition (CSD) implementation, 223–5
    - CSD echo extraction and parameter estimation, 225
    - CSD execution times and SNR for dataset of 512 samples, 226
    - resource usage and timing constraints, 225–6
    - RUSSP resource usage with split spectrum processing hardware, 226
    - split spectrum processing (SSP) implementation, 221–3
    - SSP sub-bands and results, 224
  - nondestructive evaluation and imaging applications, 205–7
- reduced instruction set CPU (RISC), 213
- reference value, 18
- resistance-to-time conversion (RTC) method, 173
- resistive gas sensors, 156
- resistive photodetectors, 156
- resistive sensor, 30–2, 154–60
  - advanced interfaces, 153–202
    - future trends, 201–2
  - bridge-type resistive sensor, 32
  - differential resistive sensor, 31–2
  - examples, 155–7
    - light dependent resistors (LDRs), 156
    - potentiometers, 157
    - resistive gas sensors, 156
    - resistive linear position based on potentiometer structure, 157
    - resistive temperature detectors (RTDs), 155–6
    - strain gauges, 156–7
  - oscillator-based systems, 175–88
    - parasitic capacitance issues, 177–82
    - problem of long measuring times, 182–88
- parallelepiped shaped usually employed to relate the electrical resistivity to resistance, 155
- parasitic capacitance, 157–60
  - effect due to sensor connections with the measurement system, 159
  - effect in gas sensors with embedded heater filament, 159
  - effect in sensors with interdigitated electrodes, 158
  - simplified model of resistive sensor taking into account the parasitic capacitive effects, 159
- resistance-to-time conversion methods, 173–200
  - basic circuit for RTC implementation with constant sensor excitation voltage, 190
  - circuit uniting, 195
  - industrial-related aspects, 199–200
  - LMS interpolation of the integrater output ramp, 192
  - modification of the circuit in fig 7.40, 197
  - problem of LMS interpolation, 192
  - sensor resistance range partition, 196
  - simple circuit for R_s estimation by using LMS estimation approach, 193
  - time diagram of circuit in fig 7.40, 197
  - time diagram of the circuit in fig 7.43, 198
  - timing diagram of circuit in fig 7.33, 189
  - timing diagram of circuit in fig 7.35, 191
- resistive sensor with single topology, differential topology and bridge topology, 31
- single resistive sensor, 31
- systems with constant sensor excitation voltage, 188–99

resistive sensor (*cont.*)
    long measuring time problem, 191–6
    parasitic capacitance estimation, 196–9
  voltamperometric resistance estimation, 160–79
    application of circuit in fig 7.10, 163
    basic circuit for RTC implementation with constant sensor, 189
    calibration function obtained with piecewise linearisation, 169
    calibration function with linear regression, 170
    calibration procedures, 167–72
    calibration results with five-decade resistance range, 172
    circuit in fig 7.7 modified to expand measurable resistance range, 163
    circuit in fig 7.9 modified to expand measurable resistance range, 164
    complete calibration function, 168
    implementation in smart sensors, 160–7
    integrator circuit, 173
    modification to circuit 7.24, 183
    modified oscillating circuit for resistance-to-time conversion, 178
    oscillating circuit for resistance-to-time conversion, 175
    oscillating circuit with moving threshold for the limitation of maximum measuring time, 186
    oscillating circuit with moving threshold with single power supply and single output signal, 187
    oscillating circuit with single power supply and single output signal, 181
    output time diagram of integrator, 174
    parasitic capacitance issues, 164–67
    resistance estimation with the voltamperometric method and known current, 160
    resistance estimation with the voltamperometric method and known voltage, 161
    resistance estimation with the voltamperometric method and known voltage applied to sensor, 162
    simulation of circuit in fig 7.10, 166
    simulation of circuit in fig7.7 when the sensor resistance has a step variation, 165
    simulation of the circuit in fig 7.8, 166
    timing diagram of circuit 7.24, 179
    timing diagram of circuit in 7.26, 181
    timing diagram of oscillator circuit, 184
    timing diagram of oscillator circuit with moving threshold approach, 184
    timing diagram of the circuit in fig 7.21, 176
    two-capacitor integrator, 185
resistive temperature detectors (RTDs), 155–6
resonant accelerometers, 364, 378–80
run card, 283

scanning electron microscope (SEM), 415–16
Schmitt trigger (ST) buffer, 39
Scopus, 21
scratch drive actuator, 324–5
  3D configuration of SDA and illustration of warping motion, 324
seismic mass, 376
self-recovery, 19
sense mode, 385
sense region, 142
sensor array
  time-of-flight measurements, 505–11
    anechoic chamber and sensor array test stand, 506
    data collected by two MEMS microphones, 508
    sample data collected by MEMS microphone array, 507

- sound source TOF estimation setup, 506
- TOF estimation experiment in an anechoic chamber, 509
- TOF estimation experiment in an anechoic chamber used 20 cycle 700 Hz sine wave sound source, 510
- TOF estimation experiment in an anechoic chamber used 700 Hz sound source, 510
- TOF estimation experiment in room environment, 508

sensor devices, 17
- interpretation of terms related to sensors, 5–10
  - about the term ‘sensor,’ 5–7
  - definitions of key terms related to devices with artificial intelligence, 7–10
- suggestions for improving terminology in the field, 17–21
  - variation of interest the specialists show the development of intelligent sensor devices, 20
- trends in development of sensors (sensor devices) and MEMS, 10–17
  - complication of organisms and sensors as tendency of evolution, 11–13
  - features and forms of intelligence, 13–17
  - method of analogy, 10–11
  - methods for providing functional reliability, 14
- what makes them ‘intelligent’ and ‘smart,’ 3–21

sensor resolution, 374

sensors, 5, 6, 17
- direct digital interface circuits, 27–59
  - applications, 53–8
  - future trends, 58–9
  - interface circuits, 40–53
  - microcontrollers, 34–40
  - sensors, 29–34

serial peripheral interface (SPI), 501

sharpness at resonance, 359

signal amplification stages, 466

signal-to-noise ratio, 389, 498

silicon, 284

silicon fusion bonding (SFB), 295

silicon-on-insulator (SOI) technology, 440, 442–3

silicon radiation detectors
- challenges for radiation detection in the EUV and VUV spectral ranges, 103–7
  - half value layer thickness of representative materials in the EUV and VUV wavelength ranges, 105
  - photodiode with external and internal photocurrent, 107
- device solutions for radiation detection in EUV and VUV spectral ranges, 107–11
  - operational scheme of pixelated diode detectors with top electrode at ground potential, 110
- future trends, 120
- methods of radiometric investigation and characterisation, 111–13
  - responsivity of ideal silicon diodes, 113
- spectral responsivity and radiation hardness of EUV and VUV radiation detectors, 113–20
  - absolute spectral responsivity of different types of photodiodes in VUV and EUV spectral ranges, 116
  - effect of local surface charging by narrow photon beams, 118
  - scale of spectral responsivity maintained by PTB, 115
  - VUV exposure of a ‘pure-B’ diode, 119
- vacuum ultraviolet and extreme ultraviolet spectral range, 102–20

silicon skeleton *see* embedded structure-based polymeric electrothermal actuators

single pole double trough (SPDT), 478

skin effect, 81

smart acoustic sensor array (SASA) system
  heart sound mapping, 512–15
    data acquired by microphones 1, 4, 11, 18, 35, 42, 49 and 52, 514
    localisation of heart valves, 513
    numbering convention for microphone array, 514
    sound imaging of heart, 515
    system collecting data on subjects heart, 513
  real-time sound processing applications, 492–515
    3D sound source localisation, 511–12
    design and implementation of smart acoustic MEMS array (AMA), 496–500
    fundamentals, 495–6
    MEMS microphones, 494
    sensor array for time-of-flight measurements, 505–11
    system implementation of AMA and CAPTAN, 500–2
  system calibration, 503–5
    frequency response of array, 505
    gain map of array and array digitally equalised, 504
    mapping array responding to stimulus of microphone, 504
  system operation, 502–3
    2D display of system acquiring data in two different situations, 503
smart communication microsystems
  radio frequency microelectromechanical systems (RF-MEMS), 472–88
    co-integration of RF-MEMS-based circuits with integrated circuits (IC), 485–7
    power capability, 483–5
    reliability, 482–3
    switches, 473
    technology and devices, 473–7
  RF-MEMS-based circuits, 478–82
    high-isolation switch, 479
    impedance tuner and true time delay phase shifter, 480
    smart microsystems illustrations, 478
    SPDT, 479
    tunable dual-frequency antenna, 481
    tunable filter, 481
smart devices
  microfabrication technologies for industrial applications, 281–303
    materials, 283–7
    MEMS design and modelling, 282–3
    processes, 287–97
    simulation, 297–302
smart energy management
  photovoltaic and fuel cells in power MEMS, 431–68
    micro-fuel cells, 449–56
    micro-fuel cells applications, 456–7
    photovoltaic mini-generators, 433–45
    photovoltaic mini-generators applications, 445–9
    power density and voltage map of different small sized power supplies, 432
  sun sensors, 457–68
    analogue sun sensor manufacturing, 463–5
    MEMS-based sun sensor, ISS-DX for industrial applications, 459
    principles and structure of sun sensor, 458–63
    strategy, 465–8
smart microelectromechanical systems
  dynamical behaviour in industrial applications, 349–64
    resonant accelerometers example, 364
  quality factor and loss of coefficient of vibrating structures, 358–63
    dependence of loss coefficients on operating time, 363
    dependence of quality factors on operating time, 363
    experimental variation of velocity and displacement in vacuum, 362

    - frequency response of microbridge resonator in ambient conditions, 361
    - frequency response of microbridge resonator in vacuum, 362
    - quality factors and loss coefficient of investigated microresonators, 360
  - resonant frequency response of vibrating structures, 350–8
    - bending modes of oscillations of electrostatically actuated MEMS cantilever, 355
    - bending modes of oscillations of electrostatically actuated MEMS microbridge, 356
    - dynamic experimental characteristics function of testing conditions, 358
    - forced vibration model with fixed support used in dynamic investigations, 352
    - frequency response of electrostatically actuated microbridge, 357
    - frequency response of electrostatically actuated microcantilever, 357
    - frequency response of electrostatically actuated microcantilever tested in vacuum, 357
    - frequency response of oscillations of electrostatically actuated microbridge tested in vacuum, 358
    - microresonators used in experimental investigations microcantilever and microbridge, 355
    - schematic of microcantilever and microbridge under electrostatic actuation, 351
- smart sensor, 7, 8
  - device, 18
  - hub, 18
- smart transducer, 7
- Snell's law, 460
- software algorithms, 463
- 'solar'-blind detectors, 107
- solid model, 298–9
  - accelerometer sensor, 299
- SONNET software, 485
- sound source localisation (3D), 511–12
  - 3D distance between receivers and transmitter, 512
  - 2D localisation geometry, 511
  - 3D localisation geometry, 511
  - experimental results for measured 3D distances, 512
- spectral irradiance, 435
- spectral response, 434–5
- spectral responsivity, 114
- specular reflection, 126
- spin casting, 289, 292
- spray deposition, 407–8
- spray micro-thrusters, 407–8
- sputter etching, 294
- sputtering, 292
- stochastic trembling, 373
- Stokes degree of linear polarisation, 135, 149
- Stokes parameters, 134, 148
- strain gauges, 156–7
- sun sensors, 457–68
  - analogue sun sensor manufacturing, 463–5
    - fabrication process of two-photodiodes, 464
    - prototype, 465
  - principles and structure of sun sensor, 458–63
    - decomposition of real sunlight beam in both x and y components, 459
    - operation principle, 460
    - refraction of light beam crossing an interface between two different materials, 461
  - strategy, 465–8
    - equivalent model of closed loop tracking system, 467
    - equivalent model of sensitive element, 466
- super-aligned carbon nanotube sheets (SACNS), 331
- super-fine inkjet printing, 424
- synchronous detection approach, 96

Talbot effect, 253
Taylor series expansion, 93
technical datasheets, 289
telesupplying bio-implantable devices, 445–7
  current consumption of commercial pacemaker with Wilson Greatbatch WG, 447
  photovoltaic area necessary to increase battery lifetime up to pacemaker consumption, 447
thermal moment, 360
thermal oxidation, 291
thermo-mechanical noise, 373–4, 393–4
thermoelastic effect, 362–3
thermoelastic theory, 360
thermoelectric cooler (TEC), 333
time difference of arrival (TDOA), 511
time-of-flight (TOF), 505–11
Time Unit block, 192
tracking system, 466–7
transducer, 5
transimpedance amplifier (TIA), 97
transmission control protocol (TCP), 209
transmittance efficiency, 140
true time delay (TTD), 479
two-beam method, 320

u-shaped beam, 326–9
  in-plane unidirectional U-shaped beams, 326
  top view of modified U-beams, 328
  unidirectional, bidirectional, and with 2DOF out-of-plane U-beams, 327
ultrasonic pulse-echo-technique, 207
universal asynchronous receiver transmitter (UART), 212
universal sensors and transducers interface (USTI), 59
User Datagram Protocol (UDP), 501
UV/visible spectral range (VIS), 114

V-beam actuator, 329–30
  schematic images of actuation, 329
vacuum ultraviolet (VUV)
  advanced silicon radiation detector, 102–20
    challenges for radiation detection in the EUV and VUV spectral ranges, 103–7
    device solutions for radiation detection in EUV and VUV spectral ranges, 107–11
    future trends, 120
    methods of radiometric investigation and characterisation, 111–13
    spectral responsivity and radiation hardness of EUV and VUV radiation detectors, 113–20
vapour phase etching, 294
Veier technique, 58
vibrometer analyser, 354
Von Mises stress, 301

weighted LMS (WLMS), 171
wet etching, 293
white noise signal, 354
WirelessHART (HEART), 202

Xilinx Embedded Development Kit, 209
Xilinx Microblaze process, 213
Xilinx Virtex-5 FPGA, 206

Young’s modulus, 330, 351

Zeeman laser, 235
Zener model, 360
‘zoom-in’ capacitor, 66